

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-208939
(P2002-208939A)

(43) 公開日 平成14年7月26日 (2002.7.26)

(51) Int.Cl.⁷

H 0 4 L 12/56

識別記号

1 0 0

F I

H 0 4 L 12/56

テ-リ-ト (参考)

A 5 K 0 3 0

1 0 0 A

審査請求 未請求 請求項の数11 O L 外国語出願 (全 67 頁)

(21) 出願番号 特願2001-357635(P2001-357635)

(22) 出願日 平成13年11月22日 (2001.11.22)

(31) 優先権主張番号 2 3 2 7 8 9 6

(32) 優先日 平成12年12月8日 (2000.12.8)

(33) 優先権主張国 カナダ (CA)

(71) 出願人 501279833

アルカテル・カナダ・インコーポレイテツ
ド

カナダ国、オンタリオ・ケー・2・ケー・
2・イー・6、カナダ、マーチ・ロード・
600

(72) 発明者 マイク・リープス

カナダ国、オンタリオ・ケー・1・エス・
3・ワイ・6、オタワ、モンク・ストリー
ト・16

(74) 代理人 100062007

弁理士 川口 義雄 (外5名)

最終頁に続く

(54) 【発明の名称】 ATMプラットフォームにおけるMPLS実装

(57) 【要約】

【課題】 通信ネットワーク内の第1のノードおよび第2のノードの間で接続パスを確立しようとする試行を計時する方法を提供すること。

【解決手段】 この方法は、それが存在する場合、以前の2回の、接続を確立しようとする試行間に前に経過した期間よりも長い期間が経過した後、接続パスを確立しようとする試行を開始する。

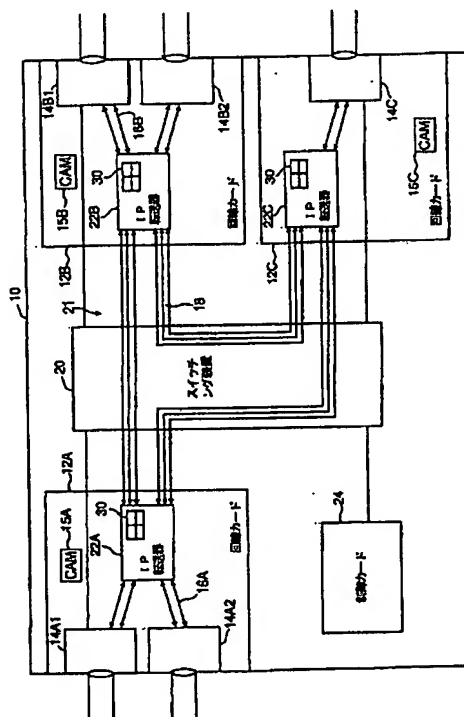


Figure 1

Best Available Copy

【特許請求の範囲】

【請求項1】 通信ネットワーク内の第1のノードと第2のノードの間で接続パスを確立しようとする試行を計時する方法であって、ある期間が経過した後に接続パスを確立しようとする試行を開始することを含み、前記期間は、前記接続を確立しようとする以前の2回の試行が存在した場合、その試行間で前に経過した別の期間よりも長い方法。

【請求項2】 前記期間が前記別の期間よりも固定時間値だけ長い請求項1に記載の方法。

【請求項3】 前記期間が最大時間値を超えない請求項1に記載の方法。

【請求項4】 前記接続パスがソフト永久ラベルスイッチパスである請求項1に記載の方法。

【請求項5】 前記固定時間値が10秒である請求項2に記載の方法。

【請求項6】 通信ネットワーク内の接続を求める複数の要求に対する接続を確立しようとする試行を計時する方法であって、一定の時間間隔の経過を追跡するタイマ構成を有することと、

接続を求める前記複数の要求に関連するレコードのリストを有することと、

前記リストから1つのレコードを選択することと、

前記1つのレコードに関連する接続を確立しようと試みることとを含み、かつ前記1つのレコードに関連する前記接続が確立された場合には、

前記レコードに成功したものとしてマークを付け、そうでなければ、前記一定の間隔で増大する連続する間隔で前記接続を確立しようと再試行することをさらに含む方法。

【請求項7】 前記リストから1つのレコードを選択することが、

レコードの前記リスト内に時間フィールドを有することと、

レコードの前記リスト内の各エントリごとに各前記一定の時間間隔で、

前記時間フィールド内の時間値を減分することと、

あるエントリに関して前記時間値がゼロである場合には、

前記エントリを前記1つのレコードとして選択することとを含む請求項6に記載の方法。

【請求項8】 連続する時間間隔で前記接続を確立しようと再試行するとき、前記連続する時間間隔は最大時間値を超えない請求項6に記載の方法。

【請求項9】 前記最大時間値が60秒である請求項8に記載の方法。

【請求項10】 2つのノード間で関連する少なくとも2つの通信リンクを有する前記2つのノードを含む通信ネットワーク内で、ラウンドロビンアルゴリズムを使用

して前記2つのノード間で通知を行うため、前記少なくとも2つの通信リンクのうちの1つを選択する方法。

【請求項11】 前記2つのノード間での通信には不十分なリソースを有する、またはそこで障害を有する前記少なくとも2つの通信リンクのうちのどの通信リンクも選択しないことをさらに含む請求項10に記載の方法。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】本発明は、デジタル通信システムに関し、より詳細には、非同期転送モード(ATM)プラットフォーム上でマルチプロトコラベルスイッチング(MPLS)を使用するネットワークノードの実装に関する。

【0002】

【従来の技術】MPLSは、インターネットプロトコル(IP)パケットを伝送するロバストな方式として、急速に業界でサポートを得ている。これは主に、MPLSではパケットのパス上のすべてのルータまたはネットワークノードで、パケットの宛先IPアドレスを検査する必要性がないからである。このため、MPLSは、多くのネットワークの高速コア内で特に有用性を有する。高速コア内にATMスイッチングインフラストラクチャが存在する可能性が高いことを認識して、業界は、現在、ATMインフラストラクチャ上でMPLSを導入するための標準を策定中である。

【0003】

【発明が解決しようとする課題】ほとんどの標準化の作業の性質と同様に、その焦点は、様々な当事者によって製造される機器の間での相互運用性を可能にするのに必要な機能上の特徴を定義することである。ただし、MPLS機能性を実装する際に多くの問題が生じる。これらには、(1)ATMインフラストラクチャ上での通知されたラベルスイッチパス(SLSP)の一般的な管理および保守、(2)SLSPの確立障害時の手続き、(3)シグナリングリンクの管理、(4)シグナリングリンクまたは物理リンクに障害が起きたときの手続き、および(5)新しいシグナリングリンクの作成などの、ネットワークボロジの様々な変更の下での手続きが含まれる。本発明は、これらの様々な問題に解決策を提供しようとするものである。

【0004】

【課題を解決するための手段】本発明の一態様は、複数の中間ノードを介してイングレスノードからイグレスノードに至る接続パスが確立される、複数の相互接続されたノードを有する通信ネットワークを管理する方法を提供する。この方法は、接続パスをネットワーク全体にわたる固有識別子と関連付けるステップと、そのパス識別子をイングレスノード上に記憶し、パスがそこから出ていることを示すようにするステップと、そのパス識別子を各中間ノード上に記憶し、そのパスがそのような各中

間ノードを通過することを示すようにするステップと、そのパス識別子をイグレスノード上に記憶し、そこにそのパスが終端することを示すようにするステップとを含む。

【0005】好ましくは、接続識別子を記憶するステップは、中間ノードを介するイングレスノードからイグレスノードに至る接続セットアップ要求の通知による接続パスの確立中に行われる。

【0006】本発明の別の態様は、初回に失敗している S L S P などの接続パスを確立しようとする試行を計時する方法に関する。これは、ある期間が経過した後、接続パスを確立しようとする新たな試行を開始することによって実現され、前記期間は、前記接続を確立しようとする試行が以前にあった場合、その 2 回の試行間で以前に経過した期間よりも長い。

【0007】本発明の別の態様は、通信ネットワークにおける、S L S P などの接続を求める複数の要求に対する接続を確立しようとする試行を計時する方法に関する。この方法は、定常時間間隔の経過を追うためのタイム構成を提供するステップと、接続を求める複数の要求に関連するレコードのリストを提供するステップと、そのリストから 1 つのレコードを選択するステップと、その 1 つのレコードに関連する接続を確立しようと試行するステップと、その 1 つのレコードに関連する接続が確立された場合、その 1 つのレコードに成功したものととしてマークを付け、そうでなければ、定常間隔ずつ長くなる間隔で連続してその接続を確立しようと再試行するステップとを含む。

【0008】その他の態様では、本発明は、前述した態様の様々な組み合わせおよびサブセットを提供する。

【0009】本発明の前述した態様およびその他の態様は、以下に記す本発明の特定実施形態の説明、および本発明の原理を例としてのみ示す添付の図面からより明白となる。図面では、同じ要素には同じ参照番号が付いており、これらの参照番号は、同じ要素の特定の例示を識別するため、固有の英字サフィックスが付いている場合がある。

【0010】

【発明の実施の形態】以下の説明およびそこでの実施形態は、本発明の原理の特定実施形態の 1 つまたは複数の例を示すものとして提供する。これらの例は、それらの原理を説明する目的で提供するものであり、それらの原理を限定するものではない。下記の説明では、本明細書および図面の全体にわたり、同じ要素には、それぞれ同じ参照番号を付けている。

【0011】1. ATMスイッチングの概要

図 1 は、例としての二重機能の ATM スイッチおよび I P ルータ 10（以降、「ノード」）のアーキテクチャを示すブロック図である。ノード 10 は、物理インターフェース入力／出力ポート 14 を有する回線カード 12 な

どの、複数の入力／出力コントローラを含む。一般的に言って、回線カード 12 は、ポート 14 で着信 ATM セルを受信する。標準化された ATM 通信プロトコルによれば、各 ATM セルは、固定サイズのものであり、また仮想パス識別子（V P I）および仮想チャネル識別子（V C I）を組み込んで、セルを特定の仮想回路（V C）と関連付けることができるようにしている。受信したそのような各セルごとに、回線カード 12 は、V C 上の対応するルックアップテーブルまたはコンテンツアドレスサブルメモリ（C A M）15 を照会する。C A M 15 は、各セルごとに、発信ポートおよびイグレス回線カードに関する事前構成されたアドレス指定情報を提供する。これは、「イグレス接続インデックス」を使用して実現され、このインデックスは、セルが次のネットワークリンク上に進む際、そのセルに帰属させられるべき新しい V C 識別子を記憶するイグレス回線カードに関する事前構成されたメモリロケーションに対するポインタである。イグレス回線カードは、アドレス指定情報およびイグレス接続インデックスを各セルに付加して、このセルをスイッチング装置 20 に送信し、このスイッチング装置は、このセルを適切なイグレス回線カードに物理的に再送する、またはコピーする。イグレス回線カードは、次に、事前構成された V P I / V C I フィールドの置き換えを行い、セルをイグレスポートから送信する。この型の ATM スwitching 機構のさらなる詳細は、P C T 公開番号 W O 9 5 / 3 0 3 1 8 で見ることができる。この公開番号はその全体が、参照により、本明細書に組み込まれる。

【0012】また、ノード 10 は、以下にずっと詳細に説明するとおり、経路指定機能およびシグナリング機能を含む様々なノード機能を制御および構成するための制御カード 24 も備えている。回線カード 12 は、スイッチング装置 20 を介して、ポート 14 で受信したデータをこの制御カード 24 に送信することができる。

【0013】各回線カードは、双方向のトラフィックフローをサポートする（すなわち、着信パケットおよび発信パケットを処理することができる）。ただし、説明では、以下の議論は、図 1 で左から右に流れるデータトラフィックに関して、回線カード 12 A およびポート 14 A 1 および 14 A 2 がイグレス処理を提供し、回線カード 12 B、12 C およびポート 14 B 1、14 B 2、14 C 1、14 C 2 がイグレス処理を提供するものと想定する。

【0014】2. I P 経路指定の概要

図示する実施形態のノードは、また、インターネットプロトコル（I P）などの階層的により高い通信レイヤと関連するデジタルデータの可変長パケットが、ATM トランスポートレイヤインフラストラクチャ上で搬送され得るようにもする。これは、各可変長パケットをトランスポートのために複数の ATM セルにセグメント化する

ことにより、可能になる。したがって、あるVCは、IPパケットを搬送することに専用にすることができ、他方、別のVCは、本来のATM通信と排他的に関連しているようにすることができる。

【0015】セルがイングレスポート14A1に着信すると、回線カード12Aは、CAM15Aにアクセスして、前述したとおり、着信セルのVCに関するコンテキスト情報を入手する。このコンテキスト情報は、VCを「サービスインターフェース」と関連付けることができる。これは、ネットワークを通るAAL5 ATMパスなどのリンクレイヤ（すなわち、「レイヤ2」）パスの端点である。各I/Oポート14上にいくつかのサービスインターフェース（SI）が存在することが可能である。これらのサービスインターフェースは、同じ回線カード上のIP転送器22で「終端する」。これは、後述するとおり、IPパケットを構成するATMセルがパケットに再組立てられ、その後、IP転送手続き（ATMスイッチング手続きに対立するものとして）が行われる。

【0016】IP転送の要点は、1つのSIで受信されたIPパケットが別のSIで送信されることである。図2に示すプロセス流れ図をさらに参照すると、IPパケットに関する転送プロセスは、ノードを介して、2つの処理段階で隔てられた3つのトランスポート段階に論理的に分割することができる。

【0017】矢印16Aによって概略で表す第1のトランスポート段階は、イングレスSIと関連するATMセルをイングレスポート14A1からイングレスIP転送器22Aに搬送する。

【0018】第2のトランスポート段階は、スイッチング装置20を介して、イングレス転送器22AからイングレスIP転送器、例えば、転送器22Bに、IPパケットを搬送する。この第2のトランスポート段階は、「接続メッシュ」21を介して実装される。この接続メッシュ内では、IP転送器22の各対の間で、8つの内部接続、つまりトランスポートインターフェース（TI）18がセットアップされる（3つのTIだけを示している）。これらのTIは、IPパケットに対する異なるサービスレベルまたはサービスクラス（COS）を可能にするように提供されている。

【0019】矢印16Bによって概略で示す第3のトランスポート段階は、イングレスIP転送器22Bからイングレスポート、例えば、ポート14B1、およびイングレスSIにIPパケットを搬送する。

【0020】第1の処理段階は、イングレスIP転送器22Aで行われ、そこで、イングレスSIと関連するATMセルがIPパケットに再組立てられる。これは、図2でステップ「A」として示している。次に、ステップ「B」で、IP転送器22Aは、ネットワーク上の「次のホップ」のための適切なイングレスSIを決定するた

め、パケットの宛先IPアドレスを検査する。この判定は、図3に概略で示すIP転送テーブル30（下記にさらに詳細に説明するのとおり、IPプロトコルから導出される）に基づく。テーブル30の各レコードは、IPアドレスフィールド32および「イングレスインターフェースインデックス」フィールド36を含む。パケットの宛先IPアドレスをIPアドレスフィールド32内でルックアップして、それに対する最長マッチ（すなわち、可能な限り遠くまでのパケットIPアドレスの宛先を解決するテーブルエントリ）を探し出す。対応するイングレスインターフェースインデックスは、実質的に、そのパケットに関するイングレス回線カード12B、イングレスIP転送器22B、およびイングレスSIを特定する（より詳細には、図8Bに関連する議論を参照）。イングレスインターフェースインデックスは、IPパケットに付加されている。

【0021】さらに、ステップ「C」で、IP転送器22Aは、パケットによってカプセル化されたサービスクラス（COS）を検査する。部分的にはカプセル化されたCOSおよび内部構成に基づいて、IP転送器22Aは、第2の段階のTI18のうちの1つを選択し、これが、所望のサービスクラスをもつイングレス転送器22Bに到達する。スイッチング装置20を通過するため、イングレスIP転送器22Aは、IPパケットをATMセル（ステップ「D」として概略で示す）に再セグメント化し、その宛先がイングレスIP転送器22Bであることを示すアドレス指定情報を各セルに付加する。

【0022】第2のより小さい処理段階は、イングレスIP転送器22Bで行われ、そこで、パケットからイングレスインターフェースインデックスが抽出され、ステップ「E」で、イングレスSIと関連するカプセル化とマッチするように変更される。したがって、イングレスSIと関連するVPI/VCIが、そのパケットに付加される。次に、パケットは、イングレスSI（「G」とラベル付けした）に対応する第3の段階のトランスポート16Bを使用して送達される。このプロセスで、パケットは、再びATMセルにセグメント化され、このセルは、イングレスSIおよび/または出力ポート14B1と関連するセル待ち行列内にバッファリングされる。また、待ち行列化が行われ、輻輳の起きる可能性のあるポイント

（「F」とラベル付けした）も、第2の処理段階と第3のトランスポート段階の間、つまり、イングレスIP転送モジュール22BとイングレスSI（「G」とラベル付けした）の間で生じる。

【0023】以上のことから、ATMプラットフォーム上でIP転送機能性を実施することは、比較的複雑なプロセスであることが理解されよう。このプロセスは、IPパケットが、（a）イングレスIP転送器22Aで再組立てされることと、（b）次に、スイッチング装置上でのトランスポートのためにセグメント化されること

と、(c) イングレス転送器22Bで再組立てされることと、(d) 次に、出力ポートからの送信のために再セグメント化されることを必要とする。さらに、イングレス転送器22Aで、トリビアルではないIPアドレスルックアップが行われなければならない。これらのステップは、各ネットワークノードで行われなければならないが、したがって、終端間通信の待ち時間を増大させる。

【0024】3. MPLS概説

すべてのパケットに対してそれぞれ、前述の手続きを行わなければならないことを回避するため、ノード10は、マルチプロトコルラベルスイッチング(MPLS)機能を提供する。従来のIP転送では、各パケットの宛先アドレスに関する最長マッチであるアドレスプレフィックスがルータのテーブル内に存在する場合、2つのパケットは同一の「転送等価クラス」(FEC)内にあるものと、通常、ルータはみなす。各ルータは、それぞれ独立にパケットを再検査し、パケットをFECに割り当てる。これとは対照的に、MPLSでは、パケットがMPLSドメインに入った際に一度だけ、パケットをFECに割り当て、そのFECを表す「ラベル」をパケットに付加する。MPLSがATMインフラストラクチャ上に導入される場合、ラベルは、特定のVC識別子である。MPLSドメイン内の後続のホップでは、IPパケットは、もはや検査されない。代わりに、ラベルが、次のホップを特定するテーブルへのインデックスおよび新しいラベルを提供する。したがって、MPLSドメイン内の後続のホップでは、パケットを構成するATMセルは、従来のATM技法を使用して交換することができる。そのようなパスは、当分野では、ラベルスイッチパス(LSP)として知られ、LSPは、ネットワークオペレータによって手作業で、永久ラベルスイッチパス(PSLP)としてセットアップされることが可能である。あるいは、ネットワークオペレータからコマンドがあった際にネットワークがパスを自動的にセットアップする、ラベル配布プロトコル(LDP)を使用することも可能である。そのようなパスは、通常、当分野では、ソフト永久LSPまたは通知されたLSP(SLSP)と呼ばれる。MPLSに関するさらなる詳細は、以下のMPLS標準案(すなわち、進行中)またはMPLS提案で見ることができ、これらはそれぞれ、参照により、本明細書に組み込まれる。

[1] E. Rosen, A. Viswanathan, R. Callon, Multiprotocol Label Switching Architecture, draft ietf-mpls-arch-06.txt.

[2] L. Andersson, P. Doolan, N. Feldman, A. Fredette, B. Thomas, LDP Specification, draft ietf-mpls-ldp-

06.txt. このLDPは、以降、「LDPプロトコル」と呼ぶ。

[3] B. Davie, J. Lawrence, K. McCloghrie, Y. Rekhter, E. Rosen, G. Swallow, P. Doolan, MPLS Using LDP and ATM VC Switching, draft ietf-mpls-atm-02.txt.

[4] B. Jamoussi, Constraint-Based LSP Setup using LDP, draft-ietf-mpls-cr-ldp-01.txt. このLDPは、以降、「CR-LDP」と呼ぶ。

[5] E. Braden他, Resource Reservation Protocol, RFC2205. このLDPは、以降、「RSVP」と呼ぶ。

【0025】ノード10は、SI連係を介してMPLS機能性を実装する。これは、管理エンティティ、つまり管理レコードのコンテキストでSIを示す図4を参照することにより、よりよく理解される。SIは、それと関連する内部ID番号を有する。ATMリンクレイヤ端点を表すことに加え、SIは、また、レイヤ3機能性に関するIPアドレスも表し、IP目的にどの型のカプセル化が使用されるかも示す。また、各SIは、いくつかの他の属性および方法と関連付けられる。特に、SIは以下の方法または適用と関連付けることができる。(1) IP転送、(2) MPLS転送、(3) IP経路指定、および(4) MPLSシグナリング。言い換えれば、ノード10は、(1) 前述のIP転送手続きを介して次のホップのルータにIPデータパケットを転送し、(2) 以下に説明するとおり、MPLS転送手続きを介してIPデータ・パケットを転送し、(3) IP経路指定プロトコルに関するメッセージを搬送するパケットを処理し、(4) MPLSシグナリングプロトコルに関するメッセージを搬送するパケットを処理するように構成することができる。

【0026】4. MPLSアーキテクチャの概要

図5は、制御カード24のハードウェアとソフトウェアのアーキテクチャをさらに詳細に示している。ハードウェアの点では、カード24は、複数の独立した物理的プロセッサ(図では長方形のボックスで表す)が関与する分散計算アーキテクチャを使用する。

【0027】プロセッサ50は、シグナリングメッセージに関するレイヤ2(「L2」)ATMアダプテーションレイヤパケットのセグメント化および再組立て機能を扱う。前述のとおり、いくつかのSIは、様々な型の経路指定プロトコルと関連付けられることになり、これらのSIのうちの1つと関連するパケットを受信した際、イングレスIP転送器22Aは、パケット(これは、スイッチング装置20を通過するために再セグメント化さ

れる)をL2プロセッサ50に送信する。再組み立てした後、L2プロセッサ50は、IP経路指定プロトコルと関連するシグナリングメッセージを経路指定プロセッサ58上で実行される「IP経路指定」68と呼ぶソフトウェアタスクに送信する。(L2プロセッサ50とIP経路指定68の間の接続は示していない。) MPLS LDPプロトコルと関連するシグナリングメッセージは、レイヤ3(L3)プロセッサ54上で実行されるラベル管理システムタスク(LMS)64に送信される。LMS64およびIP経路指定68からの発信メッセージは、適切なイグレス回線カードおよびイグレスIへのその後の送達のためにL2プロセッサ50に送信される。

【0028】IP経路指定68は、IP内部ゲートウェイ、またはI-BGP、ISIS、PIM、RIP、またはOSPFなどの経路指定プロトコルを実行する。

(これらのプロトコルに関するさらなる情報に関しては、読者は、<http://www.ietf.org/html.charters/wg-dir.html>を参照されたい。)これらの活動の結果、IP経路指定68は、図6に概略で示すマスタIP経路指定テーブル75を維持する。マスタテーブル75の各レコードは、IPアドレスのためのフィールド75a、宛先IPアドレスまたはそのプレフィックスに対応する次のホップのルータID(これは、それ自体、IPアドレスである)のためのフィールド75b、および次のホップのルータIDと関連するイグレスインターフェースインデックスのリスト75cを含む。ネットワークトポロジが変更されると、IP経路指定は、IP転送器22の転送テーブル30を適切なイグレスインターフェースインデックスをそれに送信することによって更新することになる。

(テーブル30だけが、各宛先IPアドレスエントリと関連する1つのイグレスインターフェースを有することを留意されたい。)

【0029】図5に示すとおり、ノード10は、それぞれがLMS64を含む複数のL3プロセッサ54を使用する。各LMS64は、LDPシグナリングリンクに関するTCPセッションおよびUDPセッション(LDPセッション)を終了し、各LSPに関する状態マシンを実行する。下記にさらに詳細に説明するとおり、LMS64は、LDPセッションのセットアップおよび取外しを行う要求、およびSLSPのセットアップおよび取外しを行う要求を受信する。

【0030】LMS64は、マサチューセッツ州、DedhamのHarris & Jeffriesから市販されている。相互互換性のため、ノード10は、「翻訳」ソフトウェアであるMPLSコンテキストアプリケーションマネージャ(MPLS CAM)65を含み、これは、LMSと制御カード24のそれ以外のソフトウェアエンティティの間で着信または発信要求/応答を翻

訳し、転送する。

【0031】また、各L3プロセッサ54は、呼処理タスク72も含む。このタスクは、要求された接続に関する状態情報を維持する。

【0032】別のプロセッサ56が、中央ネットワーク管理システム(NMS)(Newbridge Networks Corporation 46020(商標)製品などの)を介して、またはネットワーク端末インターフェース(NTI)を介してノードに直接に提供されるコマンド指示によって提示される管理要求を解釈し、それに応答することを含むユーザインターフェース機能性を提供する。MPLS機能性に関しては、PLSP、SLSP、およびLDPセッションをプログラムする管理要求を受け入れ、それに応答するためにユーザインターフェース66が提供される。

【0033】ノード内で接続が確立されると、リソースの割り振りおよび割り振り解除を行うためにリソース制御プロセッサ52が提供される。MPLS機能性に関しては、プロセッサ52は、LSPに対して固有ラベル値を割り当てるラベル管理タスク62を含む。

【0034】経路指定プロセッサ58上で、「MPLS経路指定」70と呼ぶソフトウェアタスクが、UI66と、IP経路指定68と、L3プロセッサ54上で実行されるLMS64との間でインターフェースをとる。一般的に言って、MPLS経路指定70はSLSPを管理する。例えば、パスセットアップ中、MPLS経路指定70は、ユーザインターフェース66からSLSPセットアップ要求を受信し、IP経路指定68から次のホップの経路指定情報を検索し、次のホップに対するLDPセッションを選択し、選択したLDPセッションを使用してSLSPパスをセットアップするためにLMS64の適切な実例を呼び出す。パスに関するラベルマッピングを受信したとき、LMS64は、MPLS経路指定70に通知を行う。次に、MPLS経路指定70が、その新しいパスのために、IP転送器22の転送テーブル30に対する更新を起動する。同様に、ネットワークトポロジが変更されたとき、MPLS経路指定70は、これらの変更をMPLS経路指定ドメイン内に反映させる。MPLS経路指定の機能が、本説明における以下の部分の焦点である。

【0035】5. 基準ネットワーク

図7は、ルータ/ノードA、B、Cの中にMPLS経路指定ドメインが存在し、ネットワーク80の残りの部分が、OSPFなどのIP特定経路指定プロトコルを使用する基準IPネットワーク80を示している。ネットワークオペレータが、ノードAから開始して、ネットワーク内のどこかに位置する宛先IPアドレス1、2、3、4(以降、「FEC Z」)に対するSLSPを確立することを所望していると想定する。(FECは、ドラフトのMPLS標準によれば、宛先IPアドレスおよびそ

のプレフィックスを含むことを留意されたい。) ネットワークオペレータは、ノードAで、そのNMTIまたはNMS (図示せず) を介して管理コマンドを入力し、FEC Zに対するSLSPの確立を要求することができる。使用するラベル配布プロトコルの型に応じて (例えば、LDPプロトコル、CRLDP、またはRSVP)、ネットワークオペレータは、SLSPに対する宛先ノードを特定する、または何らかの宛先ノードまでのSLSPに関する所望の経路を明示的に特定する (すなわち、送信元経路指定SLSP) ことさえできる。さらなる代替手法では、ラベル配布プロトコルは、FEC Zの宛先アドレスにできる限り近接したノード (MPLS経路指定ドメイン内の) を識別するのに、ベストエフォートポリシー (例えば、LDPプロトコルにおける) を使用することができる。説明する基準ネットワーク内では、ノードCが、FEC Zに対するMPLS経路指定ドメイン内の「最も近接した」ノードであると想定する。

【0036】ネットワーク80内では、ノード間でIP経路指定メッセージを通信するために、シグナリングリンク82 (これは、特定のSIと関連している) が提供される。さらに、その間でMPLSラベル配布プロトコルメッセージを通信するために、シグナリングリンク84が提供される。各シグナリングリンク84は、そのリンクと関連するLDPセッションを有する。

【0037】用語法として、文脈によってそうでないことが明らかな場合以外は、「イングレスSLSP」という用語は、発信ノード (例えば、ノードA) でのSLSPを識別するのに使用し、「通過SLSP」という用語は、通過ノード (例えば、ノードB) でのSLSPを識別するのに使用し、また「イグレスSLSP」という用語は、宛先ノード (例えば、ノードC) でのSLSPを識別するのに使用する。

【0038】図7に示す基準IPネットワークは、本発明を脈絡付けるのを助け、本発明を説明するのを助ける通常の適用例を読者に提供するために使用する。したがって、本発明は、本明細書に記載する特定の適用例によっては制限されない。

【0039】6. SLSPのデータベース管理
イングレスSLSP、通過SLSP、およびイグレスSLSPを作成し、監視し、また追跡するために、MPLS経路指定70は、図8Aのデータベース概略図に示すいくつかのテーブルまたはいくつかのデータリポジトリを維持する。各SLSPはLDPセッションによって管理されているので、各ノード上のMPLS経路指定70は、LDPシグナリングデータベース (LSLT) 100を使用して、そのノードとそのLDPピア経路指定エンティティの間でセットアップされている様々なLDPセッションを追跡する。LSLT100は、LDP経路指定ピアごとに1つのエン트리またはレコード104を

有するハッシュテーブル102を含む。レコード104は、ハッシュテーブル102に対するインデックスとして機能するルータidフィールド104a、およびLDPセッションリスト106をポイントするポインタ104b (すなわち、*ldp_session_list) を含む。ルータidフィールド104aは、1つまたは複数のLDPセッションがそれに対して構成されているLDPピアルータのIPアドレスを記憶する。各LDPセッションは、ポイントされたLDPセッションリスト106内のエン트리またはレコード108によって代表される。ノードと所与のMPLSピアルータの間で複数のLDPセッションを構成することができ、したがって、セッションリスト106は、複数のエン트리またはレコード108を有する可能性があることを留意されたい。図8Aでは、図示するルータidフィールド104aによって識別されたLDPピアルータに関して2つのLDPセッションが構成されており、したがって、それに対応するLDPセッションリスト106内に2つのレコード108が存在する。

【0040】LDPセッションリスト106の各レコード108は、下記のフィールドを含む。

【0041】ifIndex (108a) : 特定のインターフェースインデックス、およびLDPアプリケーションに関して構成されているSIを識別するノード10内の固有番号。図8Bは、このifIndexフィールドの構造をさらに詳細に示している。この構造は、SIを担う回線カード/IPモジュールに対するノード内部デバイスアドレス、イグレスポート、SI ID番号 (これは、回線カードごとにだけ固有である)、およびLDPシグナリングリンクを扱う制御カード24上のL3プロセッサ54に対する識別コードまたは内部デバイスアドレスを含む。

【0042】*fit_list_entry (108b) : FEC情報テーブル (FIT) 110に対するポインタ。下記にさらに詳細に説明するとおり、FITは、ノードから生じるすべてのイングレスSLSPを追跡する。fit_list_entryポインタ108bは、このLDPセッションと関連するイングレスSLSPのFIT110内にあるリストをポイントする。

【0043】ldp_status (108c) : ステータス指標。ステータスは、LDPセッションが使用中であるか否かを示す1ビットフラグ (図示せず)、およびそのLDPセッションのためにリソースが利用可能であるかどうかを示す1ビットフラグ (図示せず) を含む。割振りのために利用可能なラベルがないとき、または関連するSIが動作しない状態になったとき、LDPセッションには、利用可能なリソースがないとみなされる。

【0044】*next_ldp_session : 同じLDPピアルータと関連する別のLDPセッションレ

コード108に対するポインタ

FIT110は、イングレスSLSP、すなわち、そのノードから開始したSLPを追跡する。(FIT110は、通過SLSPまたはイグレスSLSPの追跡を行わないことを留意されたい。)SLSPが構成されたとき、FITエン트리またはレコード112はMPLS経路指定70によって作成され、SLSPが削除されたとき、レコード112はFIT100から除去される。

【0045】各FITエン트리またはレコード112は、以下の要素を含む。

【0046】** prev_fitEntry (112a) : 現行のエントリを参照するポインタに対するポインタ。これは、リストへの追加および除去を容易にするために使用する。

【0047】FEC : LSPに対するIP宛先。FECは、標準案に準拠して、宛先IPアドレス112bおよびLSPが宛先とするプレフィックス112cから構成される。

【0048】Srt_index (112d) : 送信元経路テーブルまたは送信元経路リスト (SRT) 114に対するインデックス。このインデックスは、LSPが送信元経路指定されない場合、値0をとり、送信元経路指定される場合、0より大きい値をとる。SLSP確立コマンドが送信元経路指定されたパスを含む場合、ルータID IPアドレスが、図示するとおり、順番にSRT114内に記憶される。

【0049】ifIndex (112e) : FECが次のホップのルータに到達するために使用されるイグレス回線カードおよびイグレスSIを特定する。このフィールドの構造は、図8Bに示したのと同じである。ただし、FIT110内で、このフィールド112eは、FECに対するイグレスデータパス (シグナリングチャネルと対立するものとして) のためのSIを特定することを留意されたい。

【0050】fecStatus (112f) : ttl値、ingressSetupフラグ、retrySeqカウンタ、およびretrySecカウンタによって表される、このFITエントリの状態 (図8Cを参照)。ttl値は、着信パケットから減分されるべき活動時間値を示す。ingressSetupフラグは、SLSPがうまく確立されたことを示す。retrySeqカウンタは、以下にさらに詳細に説明するとおり、このSLSPをMPLS経路指定がセットアップしようと試みた回数の記録をとる。retrySecカウンタは、次の再試行が試みられるまでに何秒残っているかを把握する。

【0051】lsp_id (112g) : MPLSドメイン内のSLSPを識別するのに使用する固有識別子。本実施形態では、識別子は、ノード内のLSPを一意的に識別するために、ノードIPルータIDの連結に、U

I66によって選択された固有番号を加えたものを含む。また、lsp_idは、FIT110のためのハッシュキーとして使用する。

【0052】* RWPptr (112h) : 下記にさらに詳細に説明する経路監視データベース (RWT) 120に対するポインタ。

【0053】Next. RTLPtr (112i)、prev. RTLPtr (112j) : fecStatusフィールド112fのingressSetupフラグが、対応するSLSPがうまくセットアップされなかったことを示すFITエン트리112を追跡するのに使用する順方向ポインタおよび逆方向ポインタ。これらのポインタは、FIT110内に組み込まれるretrylist (RLT) 116を実装するのに、基本的に使用する。例えば、「A」および「B」というラベルの付いたFITエン트리112が、RTL116の一部を形成する。したがって、RTLは、ノードがFIT110を迅速に検査して、すべてのピアルータに関して待ち状態のSLSPを探し出すことができるようにする。

【0054】* next_fitEntry (112k) : 現行のFEC/FITエン트리と同じLDPセッションを使用してセットアップされた次のFEC/FITエントリに対するポインタ。

【0055】RWT120は、ノードによって扱われるすべてのSLSP、すなわち、イングレス、通過およびイグレスSLSPを追跡する。RWT120は、宛先IPアドレスフィールド122a、IPプレフィックスフィールド122b、および下記にさらに詳細に説明するLSPのリスト124をポイントする* rwt-entry122cを含む。

【0056】宛先IPアドレスフィールド122aおよびプレフィックスフィールド122bは、使用する特定のラベル配布プロトコルに応じて、異なる型の管理エンティティを記憶するのに使用する。これらのエンティティは、(a) LDPプロトコルの場合、FEC、(b) 非送信元経路指定RSVPの場合、宛先ノードのルータID、(c) 厳密な送信元経路指定が行われるCR-LDPおよびRSVPの場合、次のノードのルータID、(d) ゆるい送信元経路指定が行われるCR-LDPおよびRSVPの場合、構成された送信元経路内の次のホップであることが可能である。これらはすべて、ネットワークを通してSLSPが辿る次のホップとして要約することができる。

【0057】テーブル122は、IPプレフィックスフィールド122bに基づいてハッシュされることを留意されたい。通過ノードまたはイグレスノードですべてが同じIPプレフィックスを参照する要求されたいいくつかのSLSPが存在する可能性がある。各個別SLSPは、LSPリスト124内の別々のエン트리またはレコード126によって識別される。ただし、ノード10上

10

20

30

40

50

の任意の所与のIPプレフィックスと関連するイングレスSLSPは、1つだけ存在することが可能である。

(言い換えれば、LMS64から受信された次のホップの要求ごとに1つのエントリ126が存在し、またノード上に作成された1つのイングレスSLSPに対して1つのエントリが存在する。イングレスSLSPもまた、次のホップの情報を要求し、したがって、このテーブル内に含まれることも留意されたい。)

【0058】各LSPリストエントリ126は、以下の要素を含む。

【0059】prev_RWTPtr(126a)、next_RwtPtr(126f)：特定のIPプレフィックスに関する追加のエントリ126を追跡するのに使用する順方向ポインタおよび逆方向ポインタ。同じIPプレフィックス122bと関連するLSPのすべては、ポインタ126aおよび126fを使用してまとめてリンクする。

【0060】next_EgressPtr(126b)、prev_EgressPtr(126c)：下記にさらに詳細に説明するとおり、新しいLDPセッションが構成されたとき、場合によっては拡張され得るイングレスSLSPを追跡するのに使用する順方向ポインタおよび逆方向ポインタ。これらのポインタは、基本的に、RWT120内に組み込まれるLSPイグレステーブルまたはLSPイグレスリスト(LET)130を実装するのに使用する。例えば、図8Aでは、「X」および「Y」というラベルの付いたRWTエントリ126が、LET130に属する。SLSPをセットアップする際、対応するFECの宛先アドレスに「より近接した」さらなるLDPシグナリングリンクをもはやノード10が見つけられないときにベストエフォートポリシー(すなわち、LDPプロトコル)が使用されるときはいつでも、エントリ126がLET130に追加される。例えば、基準ネットワーク内でFEC Zに対するSLSPを確立する際、ノードC(MPLS経路指定ドメインの境界にある)が、FEC Zの宛先アドレスに向うLDPシグナリングリンクを最早見つけられないとき、したがって、ノードCが、このSLSPに関するRWTエントリ126を作成するとき、エントリがLETに追加される。

【0061】fitEntryPtr(126d)：このRWTエントリ126に対応するFITエントリ112に対するポインタ。このフィールドの値は、このノードで作成されたイングレスSLSPに対するものを除き、すべてのエントリに関して零となる。

【0062】L3_id(126e)：LSPに関する次のホップの要求を最初に要求したL3プロセッサのアドレスまたは識別、またはイングレスSLSPをセットアップするのに使用するL3プロセッサのアドレスまたは識別。

【0063】lsp_id(126g)：FIT110内のlsp_id112gと同じであるが、これらのLSPは、他のノードで開始された可能性があることが異なる。

【0064】7. LDPセッションの確立

LDPセッションは、UI66を介して受信され、MPLS経路指定70に転送される管理要求を介して構成される。UI66によって得られるデータには、LDPシグナリングリンクSIのATMリンクレイヤ端点(すなわち、回線カードアドレス、ポート、VPI/VC1)、SIに割り当てられたIPアドレス、ならびにラベル範囲、ラベル空間ID、およびキープアライブタイムアウトなどのLDP特定パラメータが含まれる。

【0065】MPLS経路指定70は、ラウンドロビンアルゴリズムを使用して、LMS64の1つの実例(すなわち、L3プロセッサ54のうちの1つ)を選択し、新しいLDPセッションを確立するように関連するMPLS CAM65に要求する。MPLS CAMは、ネットワークオペレータによって選択されたSI上でLDPシグナリングアプリケーションを使用可能にし、L2プロセッサ50と関連するフィルタリング機構(図示せず)を含むノードを構成し、回線カード12と選択されたLMS/L3プロセッサ54の間で、特定のLDPシグナリングSIと関連するすべてのLDPパケットが伝搬(イングレス方向とイグレス方向の両方で)され得るようにする。これが行われると、LMS64は、該当するラベル配布プロトコル(例えば、LDPプロトコル、CRLDP、RSVP)に従って、LDPピアルータに対してLDPセッション確立メッセージを送り出す。これらのメッセージには、「ハロー」およびその他のセッション確立メッセージが含まれる。

【0066】LDPピアルータとのLDPセッションが確立されると、LMS64は、ラベルマネージャ62にLDPセッションに関するラベル交渉範囲(標準案によればLDPセッションを確立する機能の1つ)を通知する。また、LMS64は、LDPピアルータのIPアドレスをMPLS経路指定70に渡し、MPLS経路指定70は、このアドレスをLSLT100のルータIDフィールド104a内に記憶する。さらに、LMS64は、LDPシグナリングSIを識別するインターフェースインデックスをMPLS経路指定70に渡し、MPLS経路指定70は、それをLSLT100のifIndexフィールド108a内に記憶する。

【0067】8. SLSPの確立

8.1 イングレスノードでの手続き

基準ネットワークを参照すると、FEC Zに対してノードAにおいてSLSPが、ノードAのNMTIまたはノードAと通信するNMSを介してネットワークオペレータによって明示的に確立されなければならない。SLSPを構成する命令は、パラメータの1つとしてZ、す

なわち、FEC Zに関する宛先IPアドレスおよびそのプレフィックスを含む。コマンドは、UI 66によって受信されて解釈される。

【0068】UI 66は、固有のLSP IDを選択し、これは、前述のとおり、好ましくは、ノードのIPルータIDおよび固有番号の連結を含む。次に、UI 66は、FEC Zに対してSLSPを作成し、それを選択したLSP IDと関連付けるようにMPLS経路指定70に要求する。

【0069】MPLS経路指定70は、IP経路指定68からのFEC Zに関する次のホップ情報を要求する。これは、次のホップの情報を得るために非送信元経路指定LSPに関して行われ、また送信元経路指定内の情報（ネットワークオペレータによって供給される）を検証するために、送信元経路指定LSPに関しても行われる。より具体的には、MPLS経路指定70は、次の手続きを実行してこの新しいFECに関するSLSPの確立を開始する。

【0070】図9をさらに参照すると、第1ステップ150で、MPLS経路指定70が、FEC Zと同じ宛先IPアドレスおよび同じプレフィックスを有する既存のエントリ112を求めてFIT110を探索する。次にステップ152で、FIT110内にそのようなエントリが存在する場合、MPLS経路指定70は、このノードからFEC Zが既に確立されていることを示す障害コードを戻す。ステップ158で、MPLS経路指定70は、新しいFITエントリ112を作成し、それをFIT110に追加する。また、対応するエントリ126も、RWTハッシュテーブル122内のFEC Zに関するLSPリスト124内に挿入される。必要があれば、MPLS経路指定70は、FEC ZのIPプレフィックスおよびIPアドレス、または明示的経路内の第1のホップのIPプレフィックスおよびIPアドレスを含む新しいエントリ122をRWT120に追加する。

【0071】ステップ160で、MPLS経路指定70は、FEC Zに到達するための次のホップに関するピアIPアドレス（または非送信元経路指定RSVPの場合、宛先ノードのルータid、またはゆるい送信元経路指定が行われるCR-LDPおよびRSVPの場合、構成された送信元経路指定内の次のホップ）を提供するようにIP経路指定68に要求する。これが得られると、ステップ162で、MPLS経路指定70は、次のホップのルータIDに一致するLSLTエントリ102を探索する。一致するLSLTエントリが存在する場合には、ステップ164で、MPLS経路指定70は、対応するLDPセッションリスト106から利用可能なLDPセッションを選択する。これは、循環リンクされたリストであり、LSLTエントリ102内の*ldp_session_listポインタ104bが、MPLS経路指定70によって選択される次のSLSPセットア

ップのために使用されるLDPセッションをポイントするように管理される。LDPセッションが選択されると、FEC Zに関して新しく作成されたFITエントリ112が、同じLDPセッションを使用してその他のFITエントリにリンク（*prev_fitEntryポインタ112aおよび*next_fitEntryポインタ112iを介して）される。

【0072】*next_ldp_sessionポインタ108dは、LDPセッションリスト内の次のセッションをポイントする。（リスト内に1つのLDPセッションだけが存在する場合には、*next_ldp_sessionは自らをポイントする。）FIT110とLDPセッションリスト106の間のリンクが作成されると、MPLS経路指定70は、リソースを有するLDPセッションリスト内の次のセッションをポイントするように*ldp_session_listポインタ104bを更新する。これは、所与のFECに対するLDPセッションを選択することに対するラウンドロビン手法をもたらす。リソースを有するピアLDPルータに対するセッションが存在しない場合、*ldp_session_listポインタ104bは更新されない。この場合、リストは、パスがセットアップされた後、MPLS経路指定70がセッションを探すのをやめる前に、一度、検査される。

【0073】また、MPLS経路指定70が、次のホップのルータIDに一致するLSLTエントリ102を見つけれなかった場合には、対応するLDPシグナリングリンクは存在しないことも留意されたい。この場合、MPLS経路指定70は、FEC Zに関して新しく作成されたFITエントリをステップ166でRTLに追加して、ステップ168で、適切な障害コードを持って戻る。

【0074】SLSPの確立を通知するLDPセッションが選択されると、次にステップ170で、MPLS経路指定70は、SLSPのセットアップを通知するようにLMS64に要求する。ノードAのLMS64は、LDP標準案に従い、ラベル要求メッセージをその下流のLDPピアルータであるノードBに送信し、FEC Zに対するLSPのセットアップを所望することを示す。ラベル要求メッセージは、経路指定プロトコルに従い（ホップごとに、または送信元経路指定されて）MPLS経路指定ドメインを介して下流に向ってイグレスノードCまで伝播され、またラベルマッピングメッセージは、逆にイグレスノードAに戻るよう上流に伝播される。最終的に、図10に示すとおり、ラベルメッセージは、FEC Zに対してMPLS経路指定70によって選択されたLDPシグナリングリンク上で帯域内で受信されなければならない。このラベルメッセージは、ラベル、すなわち、IPパケットおよびそのATMセルをノードBに転送するのに使用すべきVPI/VCV値を識別する。

ラベルは、MPLS経路指定70およびラベルマネージャ62に渡される。さらに、ステップ174で、LMS64が、データトラフィックを扱うためにイグレス回線カードおよびポート上で使用されるSIに対するイグレスインターフェースインデックスを構成するように、呼プロセッサ72に通知する。(イグレス回線カードは、FEC Zに関するSIを通知するLDPと関連する回線カードおよびポートと同じものであることを留意されたい。)これは、FEC ZをATM VPI/VCIラベルに「バインド」する。この結付けは、MPLS経路指定70に報告され、MPLS経路指定70は、ステップ176で、FEC Zに一致するエントリ112を求めてFIT110を探索し、その時点で、呼プロセッサ72から得られたイグレスインターフェースインデックスを使用してifIndexフィールド112eが更新される。

【0075】さらに、MPLS経路指定70は、retrySeqカウンタおよびretrySecカウンタをゼロに設定することによってfecStatusフィールド112f(図8C)を更新し、またingressSetupフラグを1に設定して、セットアップが成功したことを示す。ステップ178で、MPLS経路指定70は、新しく確立されたSLSPおよびそのイグレスインターフェースインデックスに関してIP経路指定68に通知を行い、その時点で後者のタスクは、そのIP転送テーブル75(図6)を更新して、新しく確立されたイグレスインターフェースインデックス(参照番号76で概略を示す)を適切なリスト75cに追加する。IP経路指定68の方は、リスト75c内にいくつかの潜在的なイグレスインターフェースインデックスを有する可能性があり、パケットを転送するのに使用することができる。これらの選択肢の中でどれかに決めるため、IP経路指定68は、MPLSによって使用可能になったイグレスインターフェースインデックス(FECごとに1つだけ存在することが可能である)に、非MPLSイグレスインターフェースよりも高い優先順位を与える優先順位スキームを使用する。優先順位スキームは、ビットマップ75d(1つだけを示している)の機構を介して実行され、このビットマップは、イグレスインターフェースインデックスリスト75cの各エントリと関連している。ビットマップ75cは、どの型のアプリケーション、例えば、SLSPまたはIPが、イグレスインターフェースインデックスエントリと関連しているかを示す。この優先順位スキームに従って、ステップ180で、IP経路指定は、各IP転送モジュールの転送テーブル30に、新しく作成されたイグレスインターフェースインデックス76をダウンロードする。(テーブル30は、各IPアドレスまたはそのプレフィックスごとに単一のイグレスインターフェースインデックスだけをリストする。)非同期で、MPLS経路指定70も、ステ

ップ182で、FEC Zに対するイグレスSLSPがうまく作成されたことをUI66に通知する。

【0076】所定期間内にラベルマッピングメッセージが全く受信されなかった場合、またはノードBから受信されたシグナリングメッセージが、FEC Zに対するSLSPのセットアップを拒否する場合、LMS64は、ステップ184で障害をMPLS経路指定70に通知する。MPLS経路指定は、これに伴い、FEC Zに関するFITエントリ112をRTL116上に配置し、fecStatusingressSetupフィールド(図8C)をゼロに設定して、retrySeqフィールドの値を増分する(最大で6まで)。ステップ186で、MPLS経路指定は、障害をUI66に通知する。

【0077】FITエントリに関する再試行機構は、SLSPパスセットアップが、10秒、20秒、30秒、40秒、50秒、60秒で再試行されるようにする線形バックオフ機構である。毎10秒ごとに作動するMPLS経路指定70に関連する1つの再試行タイマが存在する。この時点で、MPLS経路指定は、RTL116を検査して、RTL116内の各FITエントリ112に残された時間量(図8CのretrySec)を減分する。retrySec値がゼロの場合、FITエントリ112は、RTL116から除去され、再試行順序番号が1だけ増分されて、イグレスSLSPを確立しようとする新しい試行が行われる。再試行が成功した場合、retrySeqはゼロに設定され、ingressSetupフラグは1に設定される。再試行が失敗した場合には、FITエントリは再びRTLに加えられ、retrySeqが増分される(最大順序番号は、好ましくは、6である)。retrySeqが増加されたとき、MPLS経路指定70がSLSPのセットアップを再試行する期間も、次に大きい間隔に増加される。例えば、retrySeqが2から3に増加したとき、再試行間の時間間隔は、20秒から30秒に増加する、すなわち、retrySecは30に設定される。retrySeqが6に等しいとき、再試行間の間隔は60秒である。

【0078】8.2 通過ノードでの手続き
通過ノードBで、FEC Zに関するラベル要求メッセージが、MPLSシグナリングリンク84上で受信され、L2プロセッサ50によって担当のLMS64に転送される。LMS64は、次のホップの情報をMPLS経路指定70から要求し、MPLS経路指定70の方は、FEC Zに関する次のホップのルータIDをIP経路指定68から検索して、次のホップのルータIDをRWT120内に記憶し、次のホップのLDPピアルータであるノードCに対する下流のLDPセッションを選択し、前述のとおり、このデータをLMS64に供給する。次に、LMS64は、ラベル交渉範囲(上流のノー

ドAとのLDPセッションが確立されたときに決まる)の中からVIP/VCIラベルを予約するようにラベルマネージャ62に要求する。このラベルは、ノードAに上流に向かって、ラベルマッピングメッセージが送信されるときに転送される。次に、必要な場合、上流のラベル要求メッセージを受信したLMS64は、そのラベル要求メッセージをノードCまで進めるために、下流のLDPセッションを担うLMS(異なるL3プロセッサ54上の)の別の事例を通知する。

【0079】ラベルマッピングメッセージを下流のシグナリングリンクから受信したとき、LMS64は、ラベル、すなわち、上流のノードAと関連するVP I/VCIとラベル、すなわち、下流のノードCと関連するVP I/VCIの間のクロスコネクトを確立するように呼プロセス72に通知し、下流へのデータフローを確立する。通過ノードでは、これは、前述したとおり、ATM型のクロスコネクトをもたらす。さらに、ノードAに対する上流のLDPセッションを担うLMS64は、ラベルマネージャ62によって事前に予約されたラベルとともにラベルマッピングメッセージをノードAに転送する。

【0080】送信元経路指定SLSPの場合、通過ノードBがIP経路指定70から次のホップの情報を得る必要はない可能性があることを留意されたい。ただし、これは、送信元経路指定リスト内に提供される次のホップが正確であることを通過ノードがその内部経路指定テーブルを介して確認できるようにする(例えば、次のホップが、要求された宛先IPアドレスまたはプレフィックスの下にリストされているかどうかを検査することにより)好ましい機能である。明示的に経路指定された次のホップが確認できない場合には、エラーを宣言することができる。

【0081】8. 3 イグレスノード上の手続き
イグレスノードC上では、ラベル要求メッセージが、ノードBとの上流のシグナリングリンク上で受信され、L2プロセッサ50によって担当のLMS64に転送される。LMS64は、次のホップの情報をMPLS経路指定70から要求し、MPLS経路指定70の方は、IP経路指定68から次のホップの情報を要求する。ただし、この場合、次の状況のうちの1つが生じる。(1) IP経路指定68によって戻される次のホップのルータIDは、現行のノードである、または(2)次のホップは見つかるが、次のホップまでのLDPセッションが存在しない(すなわち、MPLSドメインの端に達している)。これらの場合のどちらでも、MPLS経路指定70は、FEC Zに対するSLSPがこのノードで出なければならないことをLMS64に通知し、LMS64は、前述のとおり、上流のノードBにラベルマッピングメッセージを送信するが、FEC Zに関するラベル要求メッセージは先に進めない(またそうすることができ

ない)。この場合、MPLS経路指定70は、前述のとおり、エントリ126をRWT120に追加するが、新しく作成したRWTエントリ126もLET130に追加する。

【0082】この場合、LMS64は、IP転送のために構成されたSIを確立するように呼プロセス72に指示する。このSIは、SLSPに関してノードBとCの間でMPLSラベルとして使用されるVP I/VCIと等しいATM端点(すなわち、VP I/VCI)を有する。

【0083】9. スイッチング/経路指定活動

FEC Zに対するSLSPのセットアップは説明したので、次に、FEC Zに関連するIPパケットが処理される仕方を簡単に説明する。イグレスノードAで、IPパケットは、複数のATMセルの形式でポート14A1に着信し、これをIP転送器22Aが、構成IPパケットに再組立てする。受信パケットの宛先IPアドレスが分かると、IP転送器22Aは、「最も近接した」エントリを求めて転送テーブル30を検査する。これが、FEC Zに対するSLSPの確立に関連してIP経路指定68によってダウンロードされたFEC Zに対するエントリとなる。したがって、転送テーブル30は、イグレス回線カード12Bの識別またはアドレス、イグレスポート14B1、およびイグレスSI番号を含むイグレスインターフェースインデックス76を提供する。イグレスインターフェースインデックスは、パケットに付加される。また、イグレスIP転送器22AはTI18を選択し、部分的にはパケット内にカプセル化されたCOSフィールドに基づき、スイッチング装置20を介してパケットをイグレスIP転送器22Bに伝送する。次に、パケットは、スイッチング装置20を介する選択されたTI18上での伝送のために再セグメント化され、イグレスIP転送器22Bによって受信される。イグレスIP転送器22Bの方は、パケットに付加されたイグレスSIおよびCOS情報を抽出し、イグレスインターフェースインデックス(すなわち、イグレスSI)によって示されるカプセル化と一致するように変更する。これには、パケットにVP I/VCIラベルを付加することが含まれる。パケットは、その後ATMを構成するセルにセグメント化され、イグレスSIによって示されるVP I/VCI値とともにイグレスポート14B1から送信される。

【0084】通過ノードB上で、IPパケットに対応するATMセルがイグレスポートによって受信される。CAM15が、ATMイグレス接続インデックスに戻し、セルがATMセルとして処理される。また、イグレス回線カード12Aは、CAM15Aから検索した内部アドレス指定情報を各セルに付加し、セルをイグレス回線カードに経路指定できるようにし、このイグレス回線カードが、セルのVP I/VCI値を置き換える。次

に、イグレス回線カードは、この新しいVPI/VCI値を使用してセルを送信する。この場合、IP転送モジュール22は、スイッチング動作に積極的には関与しておらず、IPパケットを再組立てまたは再セグメント化する、またはIP経路指定ルックアップを行う必要がないことを留意されたい。

【0085】イグレスノードC上で、IPパケットに対応するATMセルが、イングレスポートによって受信され、セルによって搬送されたVPI/VCIに合せて構成されたSIに従って処理される。SIは、セルがIP転送モジュール22Aに送信され、より高いレイヤのIPパケットに再組立てが行われ、その後、通常のIPパケットとして処理されるように構成されている。

【0086】10. ネットワークトポロジ変更
10.1 新しいLDPセッション

新しいLDPセッションがノード10上で確立されたとき、LMS64は、このイベントについてMPLS経路指定70に通知し、新しいLDPセッションに関するインターフェースインデックスについて知らせる。この信号は、ノードが新しいLDPセッションの開始側であっても、応答側であっても生じる。図11の流れ図をさらに参照すると、ステップ190で、MPLS経路指定が、LSLT100内でピアルータID IPアドレスを探索する。このルータに対するLSLTエントリ194が見つかった場合には、ステップ192で、MPLS経路指定70が対応するLDPセッションリスト106を検査し、新しいLDPセッションと同じインターフェースインデックスを有するLDPセッションに対するエントリ108が存在しないことを確認する。そのようなエントリが見つからなかった場合、ステップ195で新しいエントリ108が作成される。そのようなエントリが見つかった場合、エラーが戻される。新しく構成されたLDPセッションに対するピアルータIDに一致するLSLTエントリ104が見つからなかった場合には、ステップ194で、MPLS経路指定が、新しいLSLTエントリ104を作成して、それを挿入し、その後、ステップ195でLDPセッションリストエントリ106が作成される。

【0087】ステップ196で、MPLS経路指定70が、LET130を検査する。LETに属する各RWTエントリ126に対し、対応するFECがハッシュテーブル122から決定され、ステップ200で、そのFECに対する次のホップのルータIDがIP経路指定68から要求される。ステップ201で、次のホップのルータIDが、新しく構成されたLDPセッションのピアルータIDに対して比較される。一致が見つからなかった場合、制御は、ステップ198に戻り、また一致が見つかった場合、制御は、ステップ202に移行する。ステップ202で、MPLS経路指定70は、識別されたFECのための新たに到達可能なピアルータにラベル要求

メッセージを送信するようにLMS64に指示する。

【0088】10.2 シグナリングリンク障害

ノード上でLDPセッションに障害が起きたとき、ノードは、関連するVPI/VCI範囲を使用するすべてのSLSP（ラベルマネージャ62内に記憶された）を転送するのを停止して、ノードからのクロスコネクトを除去する。また、ノードは、障害の起きたLDPセッションと関連する各SLSPの上流のピアにラベル撤回メッセージを送信する。例えば、MPLSリンク84BC

(図7)に障害が起きた場合、ノードBは、FEC Zに関するラベル撤回をイングレスノードAに送信する。ラベル撤回メッセージがイングレスノードAで受信されたとき、ノードAは、そのパスを使用することを停止し（代りに、IPホップごとの転送が使用される）、前述したステップを即時に再開してFEC Zに対するパスを再確立する。これが成功しなかった場合には、FEC Zに対するSLSPがRTL116上に配置され、その後、前述した再試行手続きが実行される。

【0089】さらに、LDPセッションが、何らかの理由で、イングレスノードA内で動作しなくなった場合、LMS64はMPLS経路指定70に知らせる。この呼の一環として、LMS64は、MPLS経路指定70にピアルータID IPアドレスを提供する。次に、MPLS経路指定70は、LSLT100のルータIDフィールド104a内でピアIPアドレスを探索する。ピアIPアドレスに対するエントリが存在しない場合、エラーが戻される。ピアIPアドレスに対するエントリ104aが存在する場合、対応するセッションリスト106で障害の起きたLDPセッションが探索される。一致するLDPセッションエントリ108が存在する場合、そのエントリがセッションリスト106から除去される。

【0090】除去されたセッションリストエントリ106の*fit_list_entryポインタ108bは、障害の起きたLDPセッションを使用するすべてのイングレスSLSPを代表するすべてのFITエントリ112のリストをポイントする。これらのエントリのそれぞれに対して、MPLS経路指定70は、前述のとおり、即時にイングレスSLSPを再確立しようと試みて、イングレスSLSPをセットアップするのに使用することができる代替のLDPセッションが存在するかを調べる。再試行が成功しなかった場合、イングレスSLSPは、RTL116上を進み、先に概略を述べた再試行手続きが行われる。

【0091】10.3 IP経路指定変更

時間の経過とともに、IP経路指定68は、FEC Zに対する新しい次のホップを発見することが可能である。例えば、基準ネットワーク内では、ノードB上のIP経路指定は、FEC Zに対する次のホップがノードD（図示せず）であるべきことを発見することができ、そのような発見をすると、ノードB上のIP経路指

定68は、FEC Zに関する新しい次のホップのルータIDをMPLS経路指定70に通知する。MPLS経路指定70は、以下のプロセスを使用して、FEC Zに対するSLSPを再経路指定する。第1に、MPLS経路指定は、IPプレフィックスアドレスに一致するRWTエントリ、例えば、IP経路指定テーブル75内で変更されたFEC Zを探索する。エントリ122が見つからなかった場合、MPLS経路指定は戻り、見つかった場合は継続して、次に、新しい次のホップDのルータIDに一致するLSLTエントリ104を探索する。LSLTエントリ104が存在する場合、したがって、新しいルータDに対するLDPセッションが存在する場合、MPLS経路指定は、ルータDに対するLDPセッションを使用して、一致するRWTエントリ122によってポイントされるRWTリスト124内の各通過SLSPを進むようにLMS64に要求する。したがって、通過SLSPは、新しい次のホップのルータDに再経路指定される。ただし、新しい次のホップのルータIDに対するLSLTエントリ102が存在しない場合、したがって、それに対するLDPセッションが存在しない場合には、MPLS経路指定70は、古いホップのルータに対応するRWTリスト124内の各通過SLSPをLET130上に配置し、そのようなSLSPをイグレスSLPとみなすべきことをLMS64に通知する。LMS64の方は、イグレスSLSPに対するイグレスSIセットアップするように呼プロセス72に指示する。

【0092】また、MPLS経路指定70は、影響を受けたFECに一致するFITエントリ112を探索する。そのFECに一致するFITエントリ112が存在し、fec_statusフィールド112fのingress_setup フラグがゼロでない（すなわち、パスがセットアップされている）場合、MPLS経路指定70は、下流のルータにラベル開放メッセージを送信することにより、LMS64がイグレスSLSPを閉じることを要求する。次に、MPLS経路指定70は、新しい次のホップに対するルータID IPアドレスに一致するLSLTエントリ104aを探索する。そのようなLSLTエントリが存在する場合には、LDPセッションが対応するLDPセッションリスト106から選択され、前述のとおり、イグレスSLSPを確立するための手続きが行われる。

【0093】10.4 物理リンク障害

2つのノード間の物理リンクに障害が起きた場合には、MPLSシグナリングとIP経路指定の両方に関するシグナリングリンク82および84（図7参照）に障害が起きる。本実施形態では、IP経路指定68は、リンクがダウンしていることに気が付き、そのリンク上のどのLDPセッションもダウンしていることにLMS64が気付く前にその経路指定テーブル75を更新する。これは、LDPセッションおよびIP経路指定内のシグナリ

ングセッションに対する「タイムアウト」期間を適切に設定して、インターフェースの障害が、MPLS経路指定70よりもずっと迅速にIP経路指定68内に反映されるようにすることによって実現される。したがって、IP経路指定68は、影響を受けたSLSPに関する新しい次のホップのルータIDについてMPLS経路指定70に通知し、また前述したとおり、MPLS経路指定70は、IP経路指定68によって識別された新しい次のホップのルータを使用して、現行のノードからこれらのSLSPパスを再経路指定する。これは、シグナリングリンクがダウンしていることをMPLS経路指定70が気付いた場合に行われるように、影響を受けたSLSPを取り外してイグレスノードに戻し、それを再通知することよりも効率的である。

【0094】以上の実施形態は、説明の目的で、ある程度、特定のなものとして記載した。本発明の趣旨および範囲を逸脱することなく、本明細書に開示する実施形態に多数の変形および変更を加えるのが可能なことが、当分野の技術者には理解されよう。

【図面の簡単な説明】

【図1】ATMセルおよびIPパケットを処理するネットワークノードのシステムを示すブロック図である。

【図2】図1のノード内でどのようにIPパケットが処理されるかを示すプロセス流れ図である。

【図3】図1のノードの入力/出力コントローラと関連するIP転送器によって使用される転送テーブルを示す図である。

【図4】図1に示すようなノードと関連する「サービスインターフェース」を表すデータ構造を示す図である。

【図5】図1のノード上の制御カードと関連するハードウェアプロセッサおよびソフトウェアプロセスのアーキテクチャを示すブロック図である。

【図6】IPネットワークと関連するマスタIP経路指定テーブルを示す図である。

【図7】IPネットワーク内のMPLSドメインを示す基準ネットワークの図である。

【図8A】通知されたラベルスイッチパス（SLSP）を管理するために図1のノードによって使用されるデータベースの概略図である。

【図8B】図8Aのデータベースのあるフィールドをより詳細に示す図である。

【図8C】図8Aのデータベースのあるフィールドをより詳細に示す図である。

【図9】SLSPを確立する際に図1のノードによって実行されるステップを示す論理流れ図である。

【図10】SLSPを確立する際に図1のノードによって実行されるステップを示す論理流れ図である。

【図11】新しいSLSPシグナリングリンクが確立された場合、ノードによって実行されるステップを示す論理流れ図である。

【符号の説明】

10 二重機能のATMスイッチおよびIPルータ
 12A、12B、12C 回線カード
 14A1、14A2、14B1、14B2、14C1、
 14C2 ポート
 15A、15B、15C CAM
 16A、16B トランスポート段階

18 トランスポートインターフェース
 20 スwitchング装置
 21 接続メッシュ
 22A、22B、22C IP転送器
 24 制御カード
 30 IP転送テーブル

【図1】

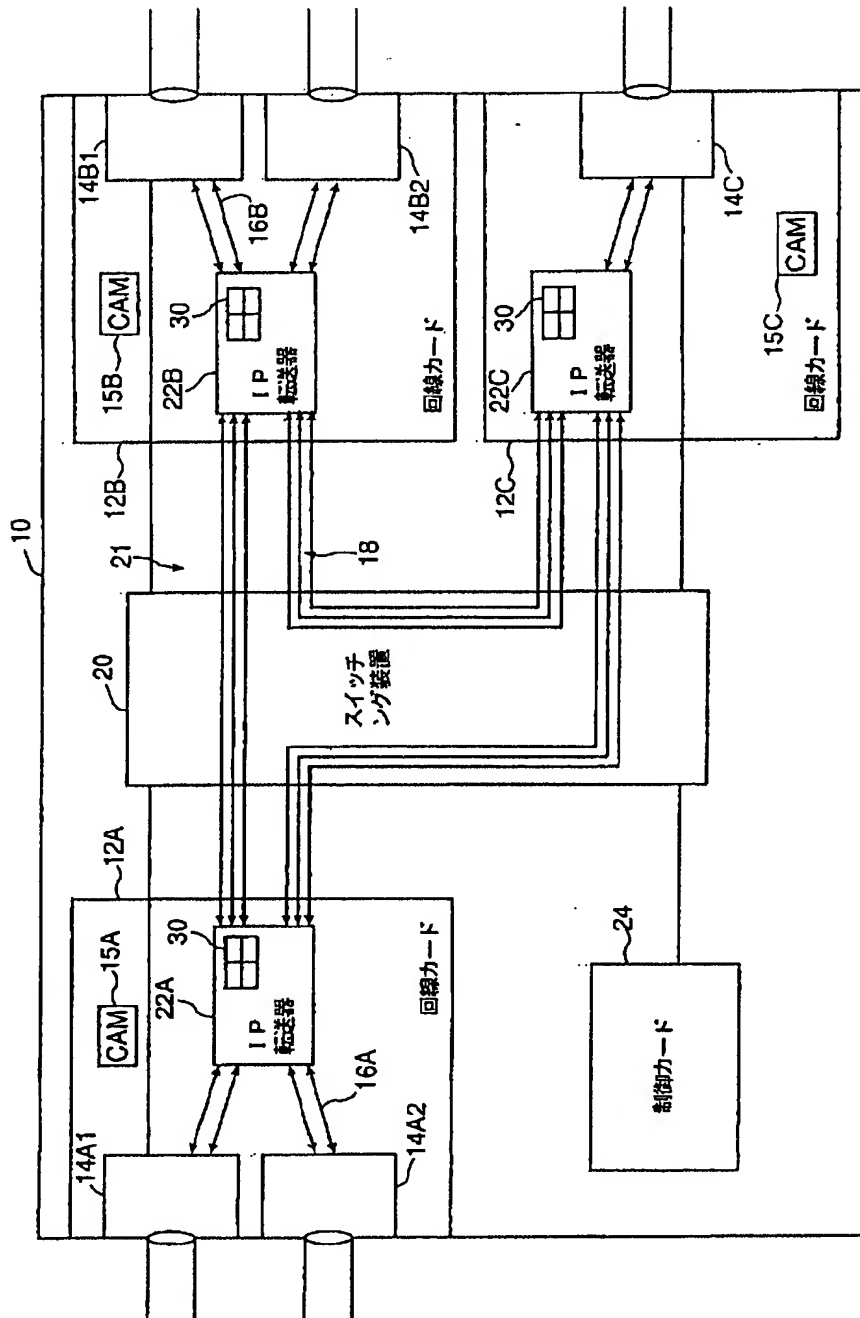


Figure 1

【図2】

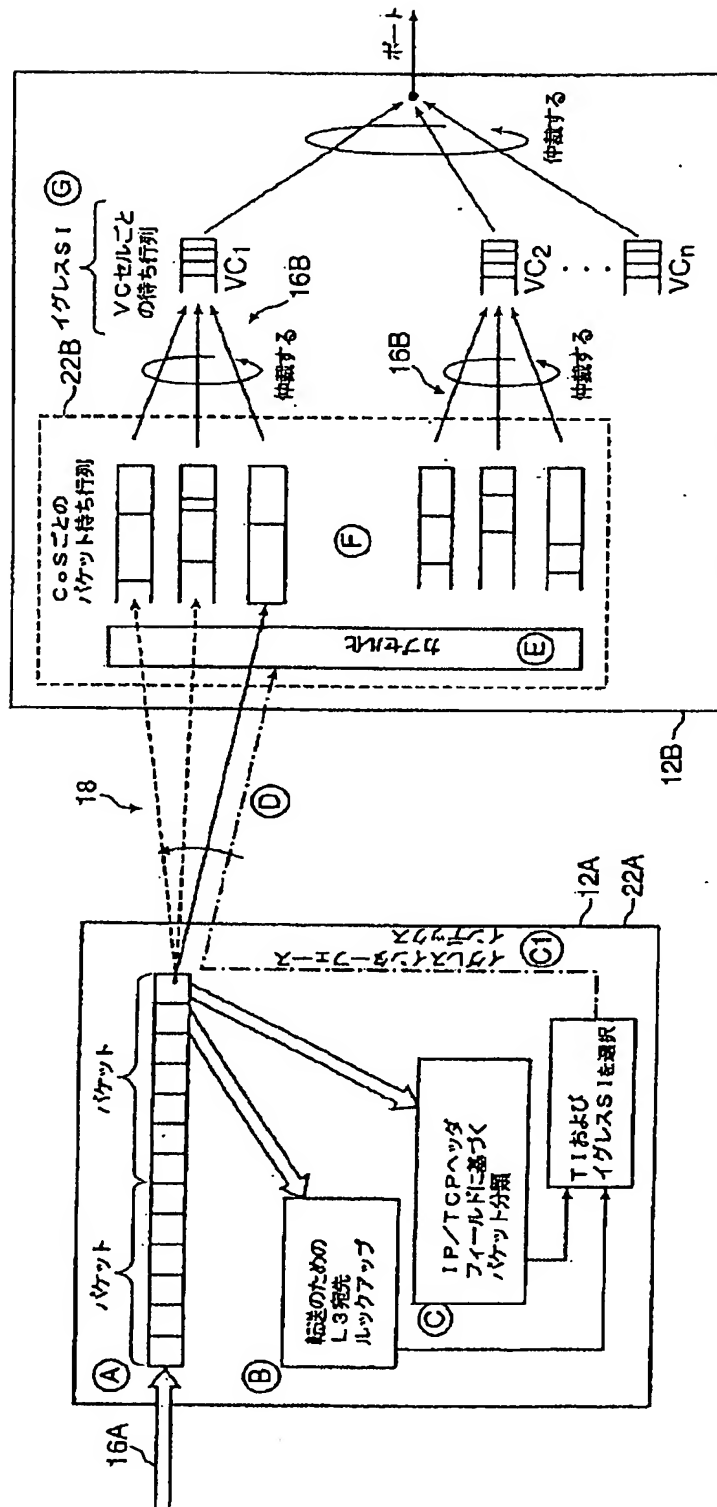


Figure 2

【図 3】

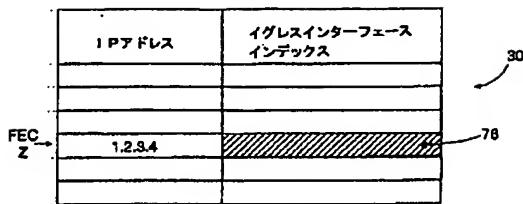


Figure 3

【図 6】

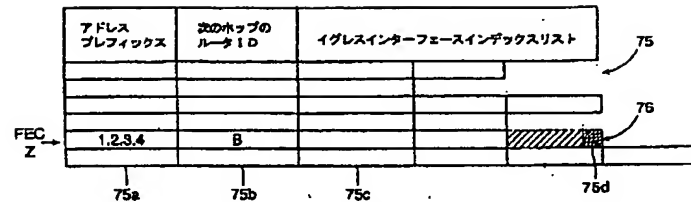


Figure 6

【図 7】

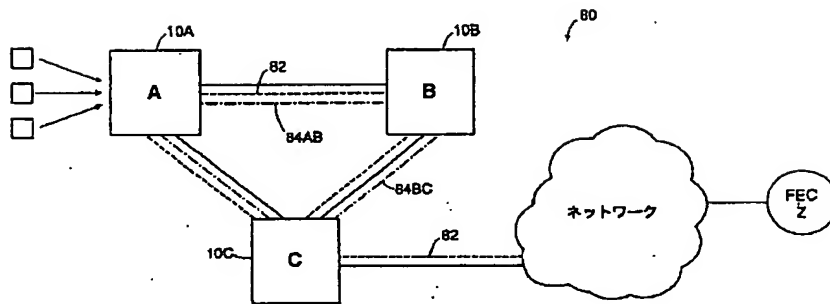


Figure 7

【図 8 B】

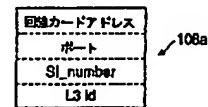


Figure 8B

【図 8 C】

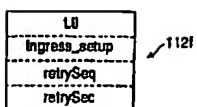


Figure 8C

Best Available Copy

【図 4】

Fig. 4

パラメータ	説明	デフォルト	NMTI	SNMP	NCI	CLI
ID番号	このS Iに割り当てられた識別番号; サブスロット内で固有。この値は、内部で割り当てられ、変更することができない。これは5桁の数のフィールドである。	なし	R	-	-	-
端点	S Iによって使用されるATM端点 (シェルフスロット-サブスロット-ポート; VP I/VC I)。	なし	R/W	-	-	-
名前	S Iの名前。これは16文字テキストストリングフィールドである。	空	R/W	-	-	-
アプリケーション	このS Iによって提供されるアプリケーション。これは、転送、経路指定、およびLDPのそれぞれが使用可能にされているかどうかを示すブールベクトル (すなわち、ビットマップ) である。	転送	R/W	-	-	-
アドレス型	IPアドレスフィールドの型。サポートされる有効な型は、無番号型およびIPv4型である。	無番号	R/W	-	-	-
IPアドレス	サービスインターフェースのIPアドレス。ユーザに対して標準の「小数点付き10進数」で表示される。「不正」IPアドレス (例えば、0. 0. 0. 0. 255. 255. 255. 255) がブロックされる。	割当てなし	R/W	-	-	-
IPアドレスプレフィックス長	(サブ) ネットワークIDを構成するIPアドレス内のビット数。0~32の範囲の数。	なし	R/W	-	-	-
隣接アドレス型	隣接IPアドレスフィールドの型。サポートされる有効な型は、無番号型およびIPv4型である。	無番号	R/W	-	-	-
隣接IPアドレス	隣接ルータにあるS Iの終端で使用するIPアドレス。ユーザに対して標準の「小数点付き10進数」で表示される。「不正」IPアドレス (例えば、0. 0. 0. 0. 255. 255. 255. 255) がブロックされる。	割当てなし	R/W	-	-	-
カプセル化	S I上で使用されるカプセル化。(RFC1483 LLC/SNAP経路指定されたIP、RFC1483 NULL)	RFC1483 NULL	R/W	-	-	-
MTU	最大伝送単位	2016 オクテット	R	-	-	-
イングレストラフィック契約	イングレストラフィック契約構造は、処置 (使用禁止、タグ、廃棄)、コミット済み情報速度 (ビット/秒での)、および bufferSize (バイトでの) から構成される。各S I内に8つのイングレストラフィック契約構造が含まれる。それぞれがCosに適用される。	使用禁止 CIR 0 BS 0	R/W	-	-	-
ステータス	サービスインターフェースのステータス。(アップ、ダウン)。	ダウン	R	-	-	-

サービスインターフェースパラメータ

Best Available Copy

【図 5】

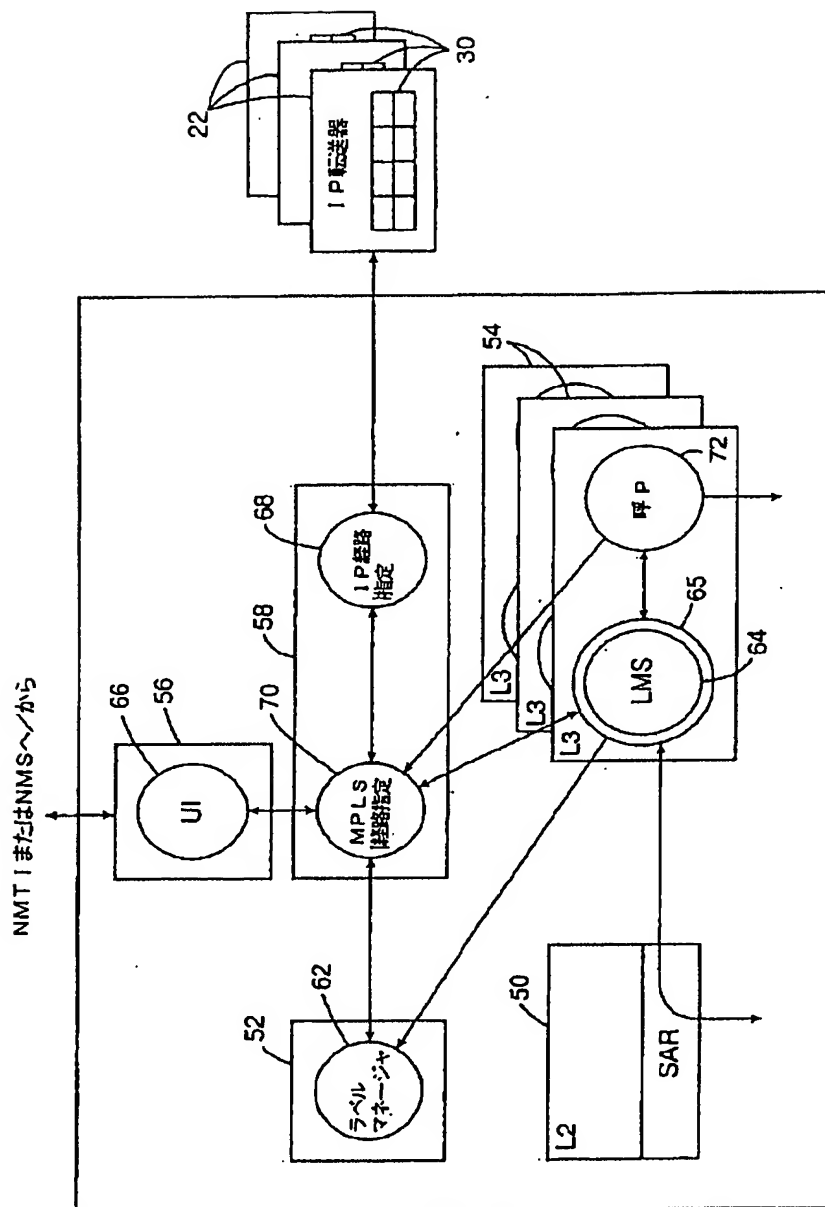


Figure 5

Best Available Copy

【図 8 A】

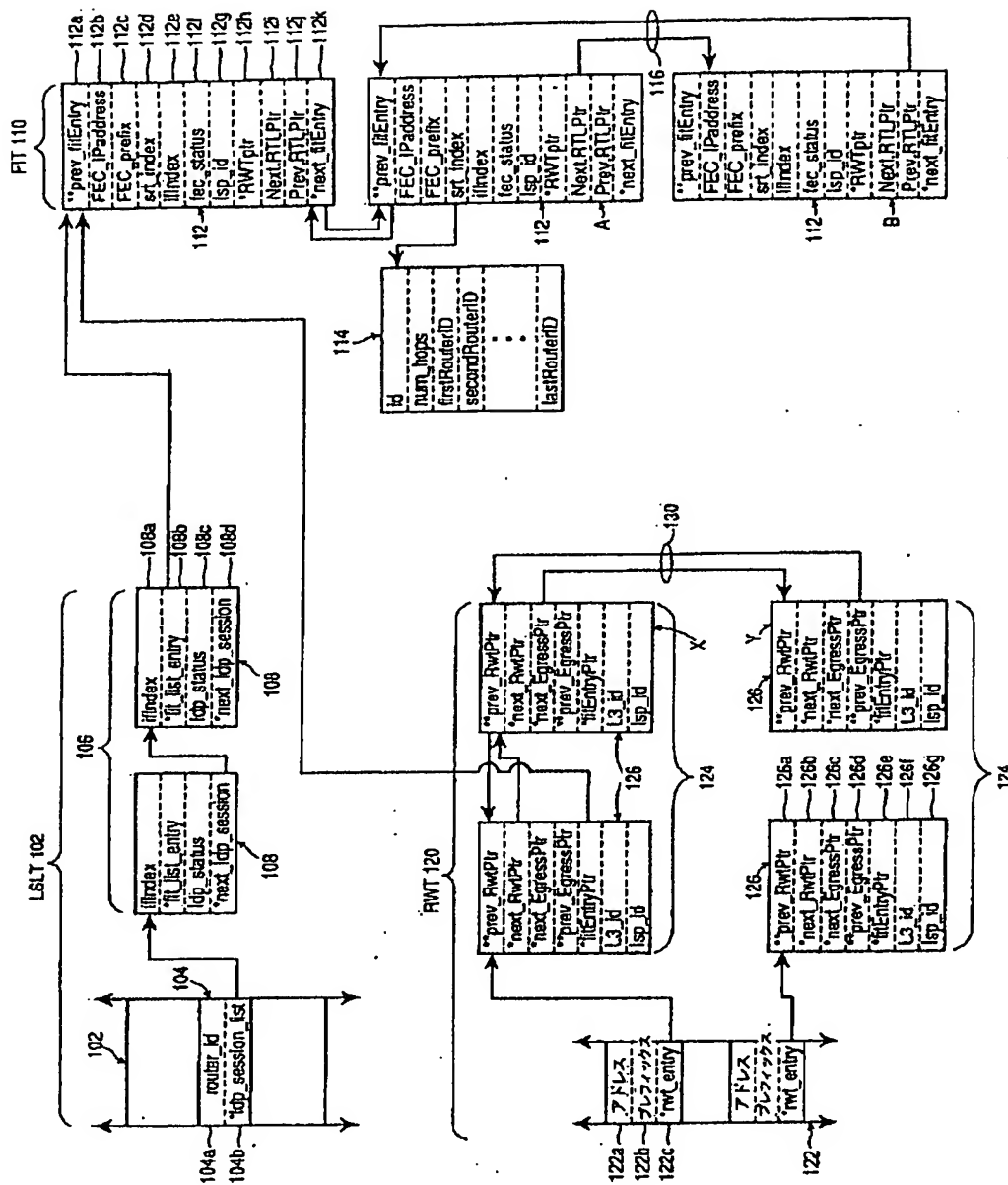


Figure 8 A

Best Available Copy

【図9】

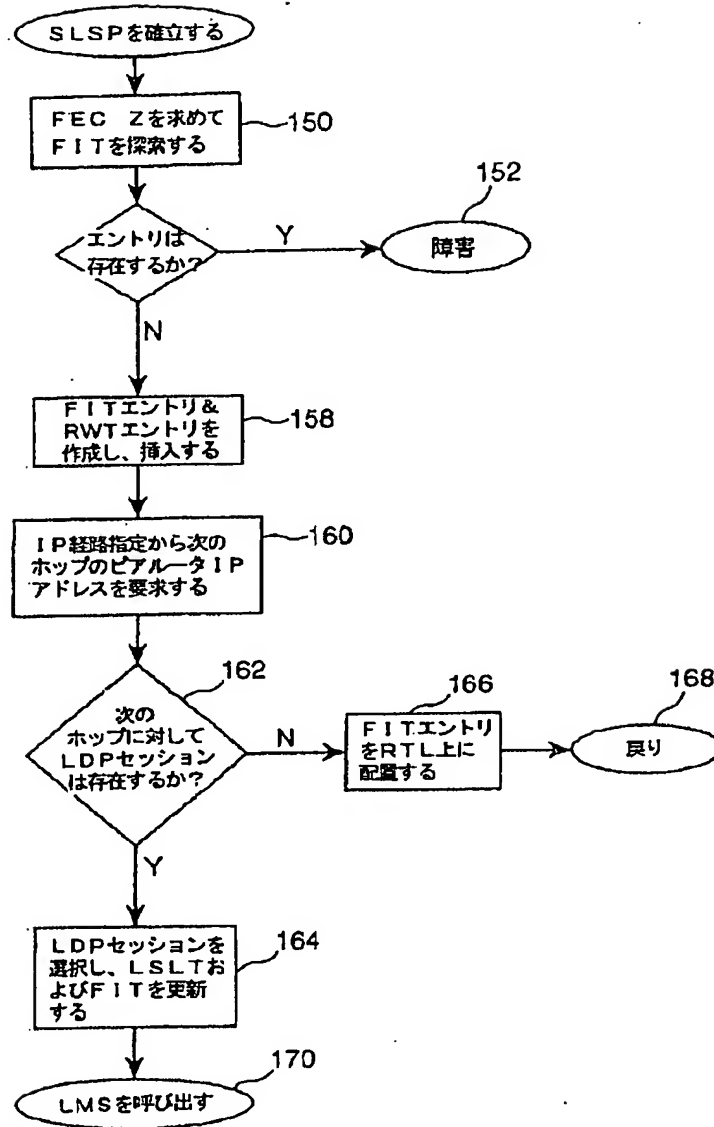


Figure 9

Best Available Copy

【図10】

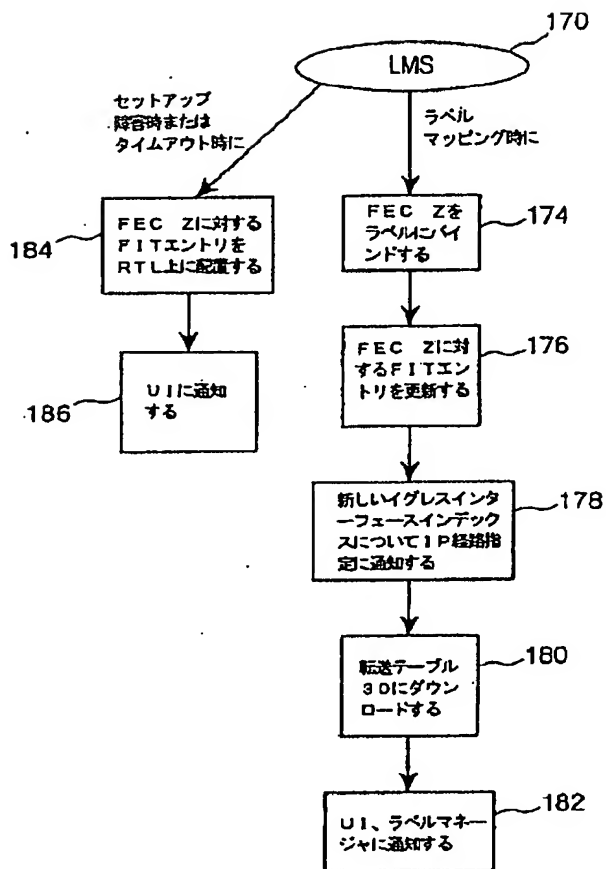


Figure 10

Best Available Copy

【図 11】

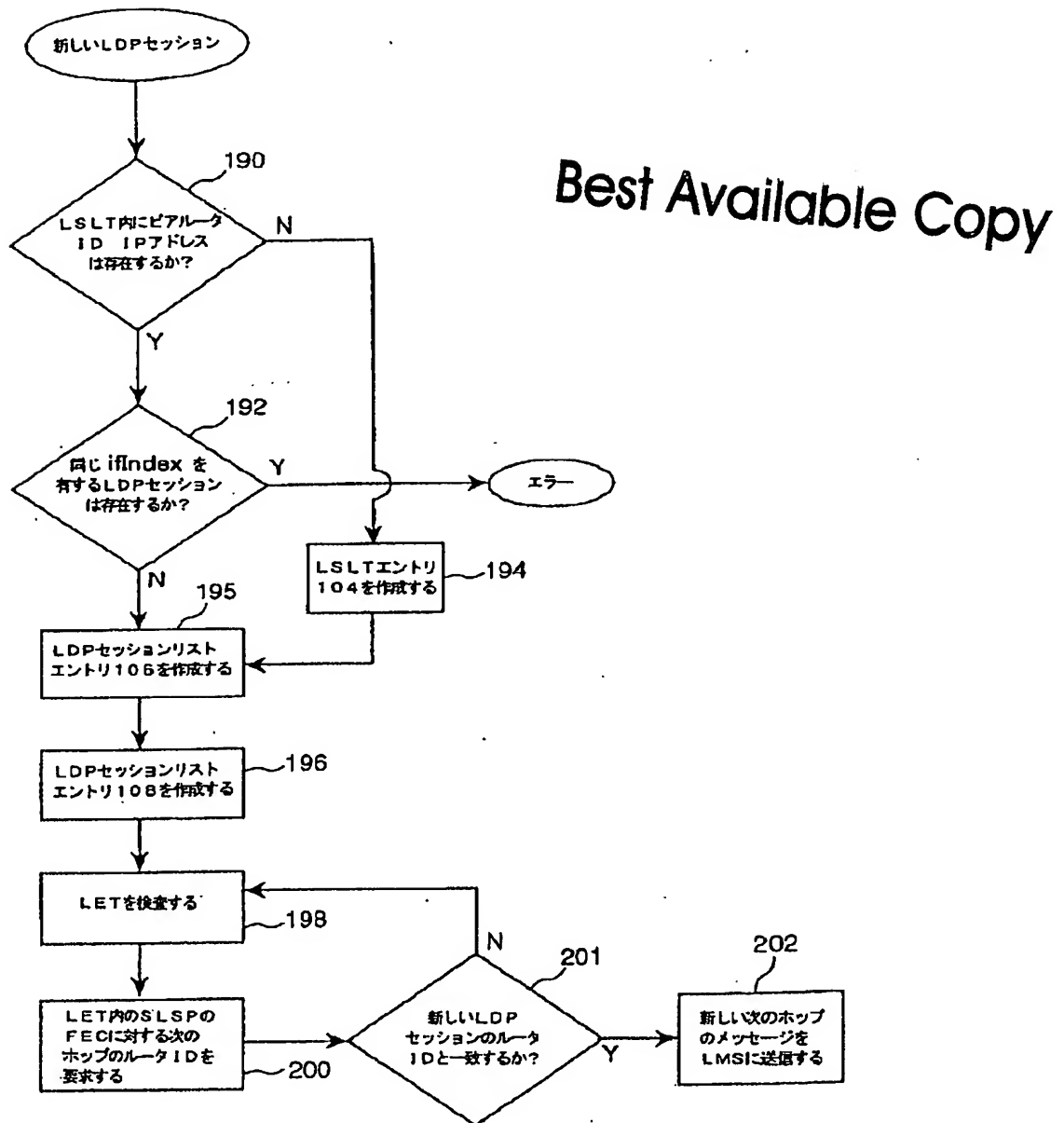


Figure 11

フロントページの続き

(72)発明者 ヌータン・ベーキ
カナダ国、オンタリオ・ケー・２・ジー・
０・エム・７、ネピーン、ムーアクロフ
ト・ロード・４０

(72)発明者 デイビッド・トスカーノ
カナダ国、オンタリオ・ケー・２・ジェ
イ・４・エヌ・９、ネピーン、バーントウ
ツド・アベニュー・２

(72)発明者 ケン・ドュブク
カナダ国、オンタリオ・ケー・２・ビー・
８・ケイ・３、オタワ、コノート・アベニ
ュー・８１１、ユニット・９１

Fターム(参考) 5K030 GA12 HA10 JL07 LB02 LB08

【外国語明細書】

1. Title of Invention

AN MPLS IMPLEMENTATION ON AN ATM PLATFORM

2. Claims

1. A method of timing an attempt to establish a connection path between a first and second node in a communications network, said method comprising initiating said attempt to establish a connection path after a period of time has elapsed wherein said period of time is greater than another period of time which had previously elapsed between two previous attempts, if any, to establish said connection.
2. The method as claimed in claim 1, wherein said period of time is greater than said another period of time by a fixed time value.
3. The method as claimed in claim 1, wherein said period of time does not exceed a maximum time value.
4. The method as claimed in claim 1 wherein said connection path is a soft permanent label switched path.
5. The method as claimed in claim 2 wherein said fixed time value is ten seconds.
6. A method of timing attempts to establish connections for a plurality of requests for connections in a communication network, said method comprising:
 - having a timer arrangement tracking passage of a regular interval of time;
 - having a list of records relating said plurality of requests for connections;
 - selecting one record from said list;
 - attempting to establish a connection relating to said one record; and
 - if said connection relating to said one record is established, then
 - marking said one record as being successful, otherwise, re-attempting to establish said connection at successive intervals increasing by said regular interval.

7. The method as claimed in claim 6 wherein said selecting one record from said list comprises:

having a time field in said list of records;

on each said regular interval of time for each entry in said list of records:

decrementing a time value in said time field; and

if said time value is zero for an entry is zero, then

selecting said entry as said one record.

8. The method as claimed in claim 6, wherein when re-attempting to establish said connection at successive time intervals, said successive time intervals do not exceed a maximum time value.

9. The method as claimed in claim 8 wherein said maximum time value is sixty seconds.

10. In a communications network comprising two nodes having at least two communications links associated between said two nodes, a method of selecting one of said at least two communications links for signalling between said two nodes utilizing a round-robin algorithm.

11. The method as claimed in claim 10 wherein said method further comprises not selecting any communications link of said at least two communications links having insufficient resources for communications between said two nodes or having a failure therein.

3. Detailed Description of Invention

FIELD OF ART

The invention relates to the art of digital communication systems and more specifically to an implementation of a network node employing multi-protocol label switching (MPLS) over an asynchronous transfer mode (ATM) platform.

BACKGROUND OF INVENTION

MPLS is quickly gaining support in the industry as a robust way of transmitting Internet Protocol (IP) packets. This is primarily because MPLS eliminates the need to examine the destination IP address of a packet at every router or network node in the path of the packet. As such, MPLS has particular utility in the high speed core of many networks. Recognizing that within the high speed core ATM switching infrastructure is likely to exist, the industry is presently in the process of formulating standards for deploying MPLS over an ATM infrastructure.

As in the nature of most standardization efforts, the focus has been to define the functional features necessary to enable interoperability amongst equipment manufactured by a variety of participants. However, many problems arise in implementing MPLS functionality. These include: (1) the general management and maintenance of signalled label switched paths (SLSP) on an ATM infrastructure; (2) procedures on the failed establishment of an SLSP; (3) management of signalling links; (4) procedures when a signalling link or physical link fails; and (5) procedures under various changes in network topology, such as the creation of a new signalling link. The present invention seeks to provide solutions to these various issues.

SUMMARY OF INVENTION

One aspect of the invention provides a method of managing a communications network having a plurality of interconnected nodes wherein a connection path is established from an ingress node to an egress node through a plurality of intermediate nodes. The method includes: associating the connection path with a network-wide unique identification; storing the path identification on the ingress node so

as to indicate that the path originates thereat; storing the path identification on each intermediate node so as to indicate that the path transits each such intermediate node; and storing the path identification on the egress node so as to indicate that the path terminates thereat.

Preferably, the steps of storing the connection identifier occurs in the process of establishing the connection path by signalling a connection set-up request from the ingress node through the intermediate nodes to the egress node.

Another aspect of the invention relates to a method of timing an attempt to establish a connection path; such as an SLSP, which has initially failed. This is accomplished by initiating another attempt to establish a connection path after a period of time has elapsed, wherein said period of time is greater than another period of time which had previously elapsed between two previous attempts, if any, to establish said connection.

Another aspect of the invention relates to method of timing attempts to establish connections for a plurality of requests for connections, such as SLSPS, in a communication network. The method includes: providing a timer arrangement for tracking passage of a regular interval of time; providing a list of records relating to the plurality of requests for connections; selecting one record from the list; attempting to establish a connection relating to the one record; and if the connection relating to the one record is established, marking the one record as being successful, otherwise, re-attempting to establish the connection at successive intervals increasing by the regular interval.

In other aspects, the invention provides various combinations and subsets of the aspects described above.

The foregoing and other aspects of the invention will become more apparent from the following description of specific embodiments thereof and the accompanying drawings which illustrate, by way of example only, the principles of the invention. In the drawings, where like elements feature like reference numerals which may bear unique alphabetical suffixes in order to identify specific instantiations of like elements),

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENT

The description which follows, and the embodiments therein, are provided by way of illustrating an example, or examples, of particular embodiments of principles of the present invention. These examples are provided for the purpose of explanation, and not limitations, of those principles. In the description which follows, like elements are marked throughout the specification and the drawings with the same respective reference numerals.

1. Overview of ATM Switching

Fig. 1 is an architectural block diagram of an exemplary dual function ATM switch and IP router 10 (hereinafter "node"). The node 10 comprises a plurality of input/output controllers such as line cards 12 which have physical interface input/output ports 14. Generally speaking, the line cards 12 receive incoming ATM cells on ports 14. Each ATM cell, in accordance with standardized ATM communication protocols, is of a fixed size and incorporates a virtual path identifier (VPI) and a virtual channel identifier (VCI) so that the cell can be associated with a particular virtual circuit (VC). For each such cell received, the line cards 12 consult a lookup table or content addressable memory (CAM) 15 keyed on VCs. The CAM 15 provides pre-configured addressing information as to the outgoing port and egress line card for each cell. This is accomplished by way of an "egress connection index", which is a pointer to a pre-configured memory location on the egress line card that stores a new VC identifier that should be attributed to the cell as it progresses its way over the next network link. The ingress line card attaches the addressing information and egress connection index to each cell and sends it to a switching fabric 20 which physically redirects or copies the cell to the appropriate egress line card. The egress line card subsequently performs the pre-configured VPI/VCI field replacement and transmits the cell out of the egress port. Further details of this type of ATM switching mechanics can be found in PCT publication no. WO95/30318, all of which is incorporated herein by reference.

The node 10 also features a control card 24 for controlling and configuring various node functions, including routing and signalling functions, as described in much

greater detail below. The line cards 12 may send data received at ports 14 to the control card 24 via the switching fabric 20.

Each line card supports bidirectional traffic flows (i.e., can process incoming and outgoing packets). However for the purposes of description the following discussion assumes that line card 12A and ports 14A1 and 14A2 provide ingress processing and line cards 12B, 12C and ports 14B1, 14B2, 14C1, 14C2 provide egress processing for data traffic flowing from left to right in Fig. 1.

2. Overview of IP Routing

The node of the illustrated embodiment also enables variable length packets of digital data associated with a hierarchically higher communications layer, such as Internet Protocol (IP), to be carried over the ATM transport layer infrastructure. This is made possible by segmenting each variable length packet into a plurality of ATM cells for transport. Certain VCs may thus be dedicated to carrying IP packets, while other VCs may be exclusively associated with native ATM communications.

When a cell arrives at ingress port 14A1 the line card 12A accesses CAM 15A to obtain context information for the VC of the arriving cell, as previously described. The context information may associate the VC with a "service interface". This is an endpoint to a link layer (i.e. "layer 2") path, such as an AAL5 ATM path, through a network. A number of service interfaces (SIs) may exist on each I/O port 14. These service interfaces "terminate" at an IP forwarder 22 on the same line card in the sense that, as subsequently described, the ATM cells constituting an IP packet are reassembled into the packet, following which IP forwarding procedures (as opposed to ATM switching procedures) are followed.

The essence of IP forwarding is that an IP packet received at one SI is retransmitted at another SI. Referring additionally to the process flow chart shown in Fig.

2, the forwarding process for IP packets can be logically divided into three transport stages, separated by two processing stages, through the node.

The first transport stage, schematically represented by arrows 16A, carries ATM cells associated with an ingress SI from the ingress port 14A1 to the ingress IP forwarder 22A.

The second transport stage carries IP packets from the ingress IP forwarder 22A across the switching fabric 20 to an egress IP forwarder, e.g., forwarder 22B. This second transport stage is implemented via a "connection mesh" 21. Within the connection mesh eight internal connections or transport interfaces (TIs) 18 are set up between each pair of IP forwarders 22 (only three TIs are shown). The TIs are provided so as to enable different levels or classes of service (COS) for IP packets.

The third transport stage, schematically represented by arrows 16B, carries IP packets from the egress IP forwarder 22B to the egress port, e.g. port 14B1, and egress SI.

The first processing stage occurs at the ingress IP forwarder 22A, where the ATM cells associated with an ingress SI are reassembled into IP packets. This is shown as step "A" in Fig. 2. At step "B" the IP forwarder 22A then examines the destination IP address of the packet in order to determine the appropriate egress SI for the "next hop" through the network. This decision is based on an IP forwarding table 30 (derived from IP protocols, as discussed in greater detail below) shown schematically in Fig. 3. Each record of table 30 includes an IP address field 32 and an "egress interface index" field 36. The IP destination address of the packet is looked up in the IP address field 32 to find the longest match thereto (i.e., the table entry which resolves the packet IP address destination as far as possible). The corresponding egress interface index essentially specifies the egress line card 12B, egress IP forwarder 22B, and the egress SI

for the packet (see more particularly the discussion with reference to Fig. 8A). The egress interface index is attached to the IP packet.

In addition, at step "C" the IP forwarder 22A examines the class of service (COS) encapsulated by the packet. Based partly on the encapsulated COS and internal configuration, the IP forwarder 22A selects one of the second-stage TIs 18 which will reach the egress forwarder 22B with a desired class of service. In order to traverse the switching fabric 20, the ingress IP forwarder 22A re-segments the IP packet into ATM cells (shown schematically as step "D") and attaches addressing information to each cell indicating that its destination is the egress IP forwarder 22B.

The second, smaller, processing stage occurs at the egress IP forwarder 22B, where the egress interface index is extracted from the packet and it is modified at step "E" to match the encapsulation associated with the egress SI. Thus, the VPI/VCI associated with the egress SI is attached to the packet. The packet is then delivered to that egress SI (labelled "G") using the third-stage transport 16B corresponding thereto. In this process the packet is segmented once again into ATM cells which are buffered in cell queues associated with the egress SI and/or output port 14B1. A queuing and possible congestion point (labelled "F") also occurs between the second processing and third transport stage — that is, between the egress IP forwarding module 22B and the egress SI (labelled "G").

It will be seen from the foregoing that effecting IP forwarding functionality on an ATM platform is a relatively involved process, requiring that the IP packet be: (a) reassembled at the ingress IP forwarder 22A, (b) subsequently segmented for transport over the switching fabric, (c) re-assembled at the egress forwarder 22B, and (d) subsequently re-segmented for transmission out of the output port. In addition, a non-trivial IP address lookup has to be performed at the ingress forwarder 22A. These steps have to be performed at each network node and hence increase the latency of end-to-end communication.

3. Introduction to MPLS

In order to avoid having to perform the above procedures on each and every packet, the node 10 provides multi-protocol label switching (MPLS) capability. In conventional IP forwarding, routers typically consider two packets to be in the same "forward equivalency class" (FEC) if there is some address prefix in that router's tables which is the longest match for the destination address of each packet. Each router independently re-examines the packet and assigns it to a FEC. In contrast, in MPLS a packet is assigned to a FEC only once as the packet enters an MPLS domain, and a "label" representing the FEC is attached to the packet. When MPLS is deployed over an ATM infrastructure, the label is a particular VC identifier. At subsequent hops within an MPLS domain the IP packet is no longer examined. Instead, the label provides an index into a table which specifies the next hop, and a new label. Thus, at subsequent hops within the MPLS domain the constituent ATM cells of a packet can be switched using conventional ATM switching techniques. Such paths are known in the art as label switched paths (LSPs), and LSPs may be manually set up as permanent label switched paths (PLSP) by network operators. Alternatively, a label distribution protocol (LDP) may be employed wherein the network automatically sets up the path upon command from the network operator. Such paths are typically referred to in the art as soft-permanent or signalled LSPs (SLSPs). Further details concerning MPLS can be found in the following draft (i.e. work in progress) MPLS standards or proposals, each of which is incorporated herein by reference:

- [1] E. Rosen, A. Viswanathan, R. Callon, *Multiprotocol Label Switching Architecture*, draft ietf-mpls-arch-06.txt.
- [2] L. Andersson, P. Doolan, N. Feldman, A. Fredette, B. Thomas, *LDP Specification*, draft-ietf-mpls-ldp-06.txt. This LDP is hereinafter referred to as "LDP Protocol".
- [3] B. Davie, J. Lawrence, K. McCloghrie, Y. Rekhter, E. Rosen, G. Swallow, P. Doolan, *MPLS Using LDP and ATM VC Switching*, draft-ietf-mpls-atm-02.txt.
- [4] B. Jamoussi, *Constraint-Based LSP Setup using LDP*, draft-ietf-mpls-cr-ldp-01.txt. This LDP is hereinafter referred to as "CRLDP".

- [5] E. Braden et al., *Resource Reservation Protocol*, RFC 2205. This LDP is hereinafter referred to as "RSVP".

The node 10 implements MPLS functionality through an SI linkage, as will be better understood by reference to Fig. 4 which shows the SI in the context of a management entity or record. The SI has an internal ID number associated therewith. In addition to representing an ATM link layer endpoint, the SI also represents an IP address for layer 3 functionality, and indicates what type of encapsulation is used for IP purposes. Each SI may also be associated with a number of other attributes and methods. In particular, SIs can be associated with the following methods or applications: (1) IP forwarding, (2) MPLS forwarding, (3) IP routing, and (4) MPLS signalling. In other words, the node 10 can be configured to (1) forward IP data packets to the next hop router via the above described IP forwarding procedures discussed above; (2) forward IP data packets via MPLS forwarding procedures as will be discussed below; (3) process packets carrying messages for IP routing protocols; and (4) process packets carrying messages for MPLS signalling protocols.

4. Overview of MPLS Architecture

Fig. 5 shows the hardware and software architecture of the control card 24 in greater detail. From the hardware perspective, the card 24 employs a distributed computing architecture involving a plurality of discrete physical processors (which are represented by rectangular boxes in the diagram).

Processor 50 handles layer 2 ("L2") ATM adaptation layer packet segmentation and reassembly functions for signalling messages. As mentioned, certain SIs will be associated with various types of routing protocols and upon receipt of a packet associated with one of these SIs the ingress IP forwarder 22A sends the packet (which is re-segmented to traverse the switching fabric 20) to the L2 processor 50. After re-assembly, the L2 processor 50 sends signalling messages associated with IP routing protocols to a software task termed "IP Routing" 68 which executes on a routing

processor 58. (The connection between the L2 processor 50 and IP Routing 68 is not shown). Signalling messages associated with MPLS LDP protocols are sent to a label management system task (LMS) 64 executing on a layer 3 (L3) processor 54. Outgoing messages from the LMS 64 and IP Routing 68 are sent to the L2 processor 50 for subsequent delivery to the appropriate egress line card and egress SI.

IP Routing 68 runs an IP interior gateway or routing protocol such as I-BGP, ISIS, PIM, RIP or OSPF. (The reader is referred to <http://www.ietf.org/html.charters/wg-dir.html> for further information concerning these protocols.) As a result of these activities IP Routing 68 maintains a master IP routing table 75 schematically shown in Fig. 6. Each record of the master table 75 includes a field 75a for an IP address field, a field 75b for the next hop router ID (which is an IP address in itself) corresponding to the destination IP address or a prefix thereof, and a list 75c of egress interface indexes associated with the next hop router ID. As the network topology changes, IP Routing will update the forwarding tables 30 of IP forwarders 22 by sending the appropriate egress interface index to it. (Note that table 30 only has one egress interface index associated with each IP destination address entry.)

As shown in Fig. 5, the node 10 employs a plurality of L3 processors 54, each of which includes an LMS 64. Each LMS 64 terminates the TCP and UDP sessions for the LDP signaling links (LDP Session) and runs a state machine for each LSP. As discussed in greater detail below, the LMS 64 receives requests to set up and tear down LDP Sessions, and to set up and tear down SLSPs.

The LMS 64 is commercially available from Harris & Jeffries of Dedham, MA. For intercompatibility purposes, the node 10 includes "translation" software, the MPLS context application manager (MPLS CAM) 65, which translates and forwards incoming or outgoing requests/responses between the LMS and the remaining software entities of the control card 24.

Each L3 processor 54 also includes a call-processing task 72. This task maintains state information about connections which have been requested.

Another processor 56 provides user interface functionality, including interpreting and replying to administrative requests presented through a central network management system (NMS) (such as the Newbridge Networks Corporation 46020™ product) or through command instructions provided directly to the node via a network terminal interface (NTI). For MPLS functionality, a user interface 66 is provided for accepting and replying to management requests to program PLSPs, SLSPs, and LDP Sessions.

A resource control processor 52 is provided for allocating and de-allocating resources as connections are established in the node. For MPLS functionality, processor 52 includes a label manager task 62 which allocates unique label values for LSPs.

On the routing processor 58, a software task termed "MPLS Routing" 70 interfaces between the UI 66, IP Routing 68 and the LMSs 64 running on the L3 processors 54. Broadly speaking, MPLS Routing 70 manages SLSPs. For example, during path setup, MPLS Routing 70 receives an SLSP setup request from the user interface 66, retrieves next hop routing information from IP Routing 68, chooses an LDP Session to the next hop, and calls the appropriate instantiation of the LMS 64 to set up the SLSP path using the selected LDP Session. When a label mapping is received for the path, the LMS 64 informs MPLS Routing 70. MPLS Routing 70 then triggers an update to the forwarding tables 30 of the IP forwarders 22 for the new path. Similarly, when the network topology changes, MPLS Routing 70 reflects these changes into the MPLS routing domain. The functions of MPLS Routing are the focus of the remainder of this description.

5. Reference Network

Fig 7 shows a reference IP network 80 wherein an MPLS routing domain exists amongst routers/nodes A, B and C, the remaining of the network 80 employing IP specific routing protocols such as OSPF. Assume the network operator wishes to establish an SLSP, commencing from node A, for IP destination address 1.2.3.4 (hereinafter "FEC Z") located somewhere in the network. (Note that a FEC, as per the draft MPLS standards, comprises a destination IP address and a prefix thereof.) The network operator may enter a management command at node A via its NMTI or the NMS (not shown) requesting the establishment of a SLSP for FEC Z. Depending on the type of label distribution protocol employed (e.g., LDP Protocol, CRLDP, or RSVP) the network operator may specify the destination node for the SLSP, or even explicitly specify the desired route for the SLSP up to some destination node (i.e., a source-routed SLSP). In the further alternative, the label distribution protocol may use a best effort policy (e.g., in LDP Protocol) to identify nodes (within the MPLS routing domain) as close as possible to the destination address of FEC Z. In the illustrated reference network, assume that node C is the "closest" node within the MPLS routing domain for FEC Z.

In the network 80, signalling links 82 (which are associated with particular SI's) are provided for communicating IP routing messages between the nodes. In addition, signalling links 84 are provided for communicating MPLS label distribution protocol messages therebetween. Each signalling link 84 has an LDP Session associated therewith.

For the purposes of nomenclature, unless the context dictates otherwise, the term "ingress SLSP" is used to identify the SLSP at the originating node (e.g., node A), the term "transit SLSP" is used to identify the SLSP at transiting nodes (e.g., node B), and the term "egress SLSP" is used to identify the SLSP at the destination node (e.g., node C).

The reference IP network shown in Fig. 7 is used to provide the reader with a typical application which will help to place the invention in context and aid in explaining it. Accordingly, the invention is not limited by the particular application described herein.

6. Database Management of SLSPs

In order to create, monitor and keep track of ingress, transit and egress SLSPs, MPLS Routing 70 maintains a number of tables or data repositories as shown in the database schema diagram of Fig. 8. Since each SLSP is managed by an LDP Session, MPLS Routing 70 on each node keeps track of the various LDP Sessions which have been set up between the node and its LDP peer routing entities using an LDP signalling database (LSLT) 100. The LSLT 100 comprises a hash table 102 having one entry or record 104 per LDP routing peer. Record 104 contains a router id field 104a which functions as the index for the hash table 102 and a pointer 104b (i.e., *ldp_session_list) which points to an LDP session list 106. The router id field 104a stores the IP address of the LDP peer router to which one or more LDP Sessions have been configured. Each LDP Session is represented by an entry or record 108 in the pointed-to LDP session list 106. Note that multiple LDP Sessions can be configured between the node and a given MPLS peer router and hence the session list 106 can have multiple entries or records 108. In Fig. 8, two LDP Sessions have been configured with respect to the LDP peer router identified by the illustrated router id field 104a, and hence two records 108 exist in the corresponding LDP session list 106.

Each record 108 of the LDP session list 106 comprises the following fields:

- ifIndex (108a) - A unique number within the node 10 which identifies a particular interface index and SI which has been configured for the LDP application. Fig. 8A shows the structure of the ifIndex field in greater detail. It comprises a node-internal device address for the line card/IP module responsible for the SI, the egress port, the SI ID number (which is only unique per line card) and an identification code or

internal device address for the L3 processor 54 on the control card 24 handling the LDP signalling link.

- ***fit_list_entry (108b)** - A pointer to a FEC information table (FIT) 110. The FIT, as described in greater detail below, keeps track of all ingress SLSPs stemming from the node. The fit_list_entry pointer 108b points to a list within FIT 110 of the ingress SLSPs associated with this LDP Session.
- **ldp_status (108c)** - A status indication. The status includes a one bit flag (not shown) indicating whether or not the LDP Session is in use and a one bit flag (not shown) indicating whether resources are available for the LDP Session. An LDP Session is considered to have no resources available when there are no labels available for allocation or when the associated SI becomes non-operational.
- ***next_ldp_session** - A pointer to another LDP Session record 108 associated with the same LDP peer router.

The FIT 110 keeps track of ingress SLSPs, i.e., SLSPs which have commenced from the node. (Note that the FIT 110 does not keep track of transit or egress SLSPs) A FIT entry or record 112 is created by MPLS Routing 70 when an SLSP is configured and record 112 is removed from the FIT 100 when an SLSP is deleted.

Each FIT entry or record 112 comprises the following elements:

- ****prev_fitEntry (112a)** - A pointer to a pointer which references the current entry. This is used for case of addition and removal from a list.
- **FEC** - IP destination for an LSP. The FEC consists of an IP destination address 112b and a prefix 112c for which the LSP is destined, as per the draft standards.

- Srt_index (112d)- An index into a source-route table or list (SRT) 114. This takes on the value 0 if the LSP is not source-routed and >0 if it is. In the event the SLSP establishment command includes a source routed path, the router ID IP addresses are stored in the SRT 114 in sequential order, as shown.
- iIndex (112e) – Specifies the egress line card and egress SI used to reach the next hop router for the FEC. The structure of this field is the same as shown in Fig. 8A. Note, however, that in the FIT 110 this field 112e specifies the SI for the egress data path (as opposed to signaling channel) for the FEC.
- fecStatus (112f) – The state of this FIT entry as represented (see Fig. 8B) by a ttl value, an ingressSetup flag, a retrySeq counter, and a retrySec counter. The ttl value indicates a time to live value that should be decremented from the incoming packets. The ingressSetup flag indicates that the SLSP is successfully established. The retrySeq counter keeps track of the number of times MPLS Routing has tried to set up this SLSP, as described in greater detail below. The retrySec counter keeps track of how many seconds are left until the next retry is attempted.
- lsp_id (112g) – A unique identifier used to identify an SLSP within an MPLS domain. In the present embodiment the identifier comprises a concatenation of the node's IP router ID plus a unique number selected by the UI 66 to uniquely identify the LSP within the node. The lsp_id is also used as a hash key for the FIT 110.
- *RWTptr (112h) – A pointer to a route watch database (RWT) 120 described in greater detail below.
- Next.RTLPtr (112i), prev.RTLPtr(112j) – Forward and backward pointers used to keep track of FIT entries 112 in which the ingressSetup flag of the fecStatus field 112f indicates that the corresponding SLSP has not been successfully set up. These pointers are basically used to implement a retry list (RTL) 116 which is embedded in

the FIT 110. For example, the FIT entries 112 labelled "A" and "B" form part of the RTL 116. The RTL thus enables the node to quickly traverse the FIT 110 to find pending SLSPs for all peer routers.

- *next_fitEntry (112k) – A pointer to the next FEC/FIT entry which has been set up using the same LDP Session as the current FEC/ FIT entry.

The RWT 120 keeps track of all SLSPs handled by the node, i.e., ingress, transit and egress SLSPs. The RWT 120 comprises a hash table 122 which includes an IP designation address field 122a, an IP prefix field 122b, and a *rwt-entry 122C which points to a list 124 of LSPs described in greater detail below.

The IP destination address and prefix fields 122a and 122b are used to store different types of management entities depending on the particular label distribution protocol employed. These entities may be: (a) the FEC, for LDP Protocol; (b) the destination node's router ID, for non source-routed RSVP; (c) the next node's router ID for strict source-routed CR-LDP and RSVP; and (d) the next hop in the configured source-route for loose source-routed CR-LDP and RSVP. These can all be summarized as the next hop that an SLSP takes through the network.

Note that table 122 is hashed based on the IP prefix field 122b. There can be several requested SLSPs all referring to the same IP prefix at a transit node or egress node. Each individual SLSP is identified by a separate entry or record 126 in the LSP list 124. However, there can only be one ingress SLSP associated with any given IP prefix on the node 10. (In other words, an entry 126 exists for every next hop request received from the LMS 64 as well as one entry for an ingress SLSP which has been created on the node. Note too that egress SLSPs also request next hop information and therefore are included within this table).

Each LSP list entry 126 comprises the following elements:

- prev_RwtPtr (126a), next_RwtPtr (126f) – Forward and backward pointers used to keep track of additional entries 126 for a specific IP prefix. All of the LSPs associated with the same IP prefix 122b are linked together using pointers 126a and 126f.
- next_EgressPtr (126b), prev_EgressPtr (126c) – Forward and backward pointers used to keep track of egress SLSPs which may possibly be extended when a new LDP Session is configured, as discussed in greater detail below. These pointers are basically used to implement an LSP egress table or list (LET) 130 which is embedded in the RWT 120. For example, in Fig. 8 the RWT entries 126 labelled "X" and "Y" belong to the LET 130. An entry 126 is added to the LET 130 whenever a best effort routing policy (e.g., LDP Protocol) is employed in setting up an SLSP and the node 10 can find no further LDP signalling links "closer" to the destination address of the corresponding FEC. For example, in establishing an SLSP for FEC Z in the reference network, node C (which lies at the boundary of the MPLS routing domain) cannot find any more LDP signalling links heading towards the destination address of FEC Z, and thus when node C creates a RWT entry 126 for this SLSP the entry will be added to the LET.
- fitEntryPtr (126d) – Pointer to the FIT entry 112 which corresponds to this RWT entry 126. The value of this field will be null for all entries except for ingress SLSPs created at this node.
- L3_id (126e) – The address or identity of the L3 processor which initially requested the next hop request for the LSP or the address or identity of the L3 processor which is used to set up an ingress SLSP.
- lsp_id (126g) – Same as lsp_id 112g in FIT 110, except that these LSPs may have been initiated at other nodes.

7. Establishing an LDP Session

LDP Sessions are configured via management requests which are received through the UI 66 and forwarded to MPLS Routing 70. The data obtained by the UI 66 includes the ATM link layer end point of the LDP signalling link SI (i.e. – line card address, port, VPI/VCI), IP address assigned to the SI, and LDP specific parameters such as label range, label space ID and keep-alive timeout.

MPLS Routing 70 employs a round-robin algorithm to select one instantiation of the LMS 64 (i.e., one of the L3 processors 54) and requests the associated MPLS CAM 65 to establish a new LDP Session. The MPLS CAM enables the LDP signalling application on the SI selected by the network operator and configures the node, including a filtering mechanism (not shown) associated with the L2 processor 50, to allow all LDP packets associated with a particular LDP signalling SI to be propagated (in both the ingress and egress directions) between the line cards 12 and the selected LMS/L3 processor 54. Once this is carried out, the LMS 64 sends out LDP session establishment messages to the LDP peer router in accordance with the applicable label distribution protocol (e.g., LDP Protocol, CRLDP, RSVP). These include “hello” and other session establishment messages.

Once an LDP Session has been established with the LDP peer router, the LMS 64 informs the label manager 62 of the negotiated label range for the LDP Session (which is a function of establishing an LDP Session as per the draft standards). The LMS 64 also passes the IP address of the LDP peer router to MPLS Routing 70 which stores this address in the router ID field 104a of the LSLT100. In addition, the LMS 64 passes the interface index identifying the LDP signalling SI to MPLS Routing 70 which stores it in the ifIndex field 108a of the LSLT 100.

8. Establishing an SLSP

8.1 *Procedures at the Ingress Node*

Referring to the reference network, an SLSP must be explicitly established at node A for FEC Z by the network operator via the NMT1 of node A or via the NMS which communicates with node A. The instruction to configure the SLSP includes as one of its parameters Z, i.e., the destination IP address and prefix thereof for FEC Z. The command is received and interpreted by the UI 66.

The UI 66 selects a unique LSP ID which, as previously discussed, preferably comprises a concatenation of the node's IP router ID and a unique number. The UI 66 then requests MPLS Routing 70 to create an SLSP for FEC Z and associate it with the selected LSP ID.

MPLS Routing 70 requests next hop information for FEC Z from IP Routing 68. This will occur for non-source-routed LSPs in order to obtain the next-hop information as well as for source-routed LSPs in order to verify the information in the source-route (which will be supplied by the network operator). More specifically, MPLS Routing 70 executes the following procedure to initiate the establishment of an SLSP for this new FEC.

Referring additionally to Fig. 9, at a first step 150 MPLS Routing 70 searches the FIT 110 for an existing entry 112 having the same IP destination address and prefix as FEC Z. If such an entry exists in the FIT 110 then at step 152 MPLS Routing 70 returns with a failure code indicating that FEC Z has already been established from this node. At step 158, MPLS Routing 70 creates a new FIT entry 112 and appends it to the FIT 110. A corresponding entry 126 is also inserted into the LSP list 124 for FEC Z in the RWT hash table 122. If necessary, MPLS Routing 70 adds a new entry 122 to the RWT 120 which includes the IP prefix and address of FEC Z, or the IP prefix and address of the first hop in the explicit route.

At step 160 MPLS Routing 70 requests IP Routing 68 to provide the peer IP address for the next hop to reach FEC Z (or the destination node's router id for non source-routed RSVP, or the next hop in the configured source-route for loose source-routed CR-LDP and RSVP). Once obtained, at step 162 MPLS Routing 70 searches for an LSLT entry 102 which matches the next hop router ID. If a matching LSLT entry exists, then at step 164 MPLS Routing 70 selects an available LDP Session from the corresponding LDP Session list 106. This is a circular linked list, which is managed such that the *ldp_session_list pointer 104b in the LSLT entry 102 points to the LDP Session to be used for the next SLSP setup which is selected by MPLS Routing 70. Once the LDP Session is selected, the recently created FIT entry 112 for FEC Z is linked (via the **prev_fitEntry and *next-FitEntry pointers 112a and 112i) to other FIT entries using the same LDP Session.

The *next_ldp_session pointer 108d points to the next session in the LDP session list. (If there is only one LDP Session in the list then the *next_ldp_session points to itself.) Once the link between the FIT 110 and LDP session list 106 is created, MPLS Routing 70 updates the *ldp_session_list pointer 104b to point to the next session in the LDP session list with resources. This results in a round robin approach to selecting LDP Sessions for a given FEC. If no sessions to the peer LDP router have resources, the ldp_session_list pointer 104b is not updated. In this case, the list is traversed once when a path is setup before MPLS Routing 70 stops looking for a session.

Note also that if MPLS Routing 70 does not find an LSLT entry 102 which matches the next hop router ID, then no LDP signaling link exists thereto. In this case MPLS Routing 70 adds the recently created FIT entry for FEC Z to the RTL at step 166 and returns at step 168 with an appropriate failure code.

Once an LDP Session has been selected to signal the establishment of the SLSP, then at step 170 MPLS Routing 70 requests the LMS 64 to signal the set up of an SLSP. The LMS 64 of node A sends a label request message, as per the draft LDP

standards, to its downstream LDP peer router, node B, indicating the desire to set up an LSP for FEC Z. The label request message is propagated downstream across the MPLS routing domain in accordance with the routing protocol (hop-by-hop or source routed) to the egress node C, and label mapping messages are propagated upstream back to the ingress node A. Ultimately, as shown in Fig. 10, a label message should be received inbound on the LDP signalling link selected by MPLS Routing 70 for FEC Z. This label message identifies the label, i.e., VPI/VCI value, that should be used to forward IP packets and the ATM cells thereof to node B. The label is passed to MPLS Routing 70 and to the label manager 62. In addition, at step 174 the LMS 64 signals the call processor 72 to configure an egress interface index for the SI being used on the egress line card and port to handle the data traffic. (Note that the egress line card will be the same line card and port associated with the LDP signaling SI for FEC Z.) This "binds" FEC Z to the ATM VPI/VCI label. The binding is reported to MPLS Routing 70 which searches the FIT 110 at step 176 for the entry 112 matching FEC Z, whereupon the ifIndex field 112e is updated with the egress interface index obtained from the call processor 72.

In addition, MPLS Routing 70 updates the fecStatus field 112f (Fig. 8B) by setting the retrySeq and retrySec counters to zero and sets the ingressSetup flag to one thereby indicating successful set up. At step 178 MPLS Routing 70 informs IP Routing 68 about the newly established SLSP and its egress interface index whereupon the latter task updates its IP forwarding table 75 (Fig. 6) to add the newly established egress interface index (shown schematically by ref. no. 76) to the appropriate list 75c. IP Routing 68, in turn, may have a number of potential egress interface indexes in list 75c, which may be used to forward a packet. In order to decide amongst these alternatives, IP Routing 68 employs a priority scheme which grants an MPLS-enabled egress interface index (there can only be one per FEC) higher priority than non-MPLS egress interfaces. The priority scheme is carried out through the mechanism of a bit map 75d (only one shown) which is associated with each entry of the egress interface index list 75c. The bit map 75c indicates what type of application, e.g., SLSP or IP, is associated with the egress interface index entry. Following this priority scheme, at step 180 IP Routing downloads

the newly created egress interface index 76 to the forwarding tables 30 of each IP forwarding module. (Table 30 only lists a single egress interface index for each IP address or prefix thereof). Asynchronously, MPLS Routing 70 also informs the UI 66 at step 182 that the ingress SLSP for FEC Z has been successfully created.

In the event that no label mapping message is received within a predetermined time period, or the signalling message that is received from node B denies the setup of an SLSP for FEC Z, the LMS 64 informs MPLS Routing 70 of the failure at step 184. MPLS Routing consequently places the FIT entry 112 for FEC Z on the RTL 116, sets the fecStatus ingressSetup field (Fig. 8B) to zero and increments the value of the retrySeq field (up to a max of 6). At step 186, MPLS Routing informs the UI 66 of the failure.

The retry mechanism for FIT entries is a linear back off mechanism which causes an SLSP path setup to be retried at 10, 20, 30, 40, 50, and 60 seconds. There is one retry timer associated with MPLS Routing 70 which goes off every 10 seconds. At this point MPLS Routing traverses the RTL 116, decrementing the amount of time (retrySec – Fig.8B) left for each FIT entry 112 in the RTL 116. If the retrySec value is zero, the FIT entry 112 is removed from the RTL 116, the retry sequence number is incremented by one and another attempt is made to establish the ingress SLSP. If the retry is successful retrySeq is set to zero and the ingressSetup flag is set to 1. If the retry is unsuccessful then the FIT entry is added back to the RTL, retrySeq is incremented (max. sequence number is preferably 6). When the retrySeq counter is increased, the time period within which MPLS Routing 70 will retry to set up the SLSP also increases to the next highest interval. For instance, when retrySeq increases from 2 to 3 the time interval between retries increases from 20 to 30 seconds, i.e. retrySec is set to 30. When retrySeq is equal to 6, retries are 60 seconds apart.

8.2 Procedures at Transit Nodes

At transit node B, a label request message for FEC Z is received on MPLS signaling link S4 and forwarded by the L2 processor 50 to the responsible LMS 64. The LMS 64 requests next hop information from MPLS Routing 70 which, in turn, retrieves the next hop router ID for FEC Z from IP Routing 68, stores the next hop router ID in the RWT 120, selects a downstream LDP Session to the next hop LDP peer router, node C, and supplies this data to the LMS 64, as discussed previously. The LMS 64 then requests the label manager 62 to reserve a VPI/VCI label from within the negotiated label range (determined when the LDP Session with the upstream node A was established). This label is forwarded upstream to node A when the label mapping message is sent thereto. Then, if necessary, the LMS 64 which received the upstream label request message will signal another instantiation of the LMS (on a different L3 processor 54) responsible for the downstream LDP Session in order to progress the Label Request message to node C.

When a label mapping message is received from the downstream signalling link, the LMS 64 signals the call processor 72 to establish a cross-connect between the label, i.e., VPI/VCI, associated with upstream node A and the label, i.e., VPI/VCI, associated with the downstream node C to thereby establish downstream data flow. On the transit node this results in an ATM style cross-connect, as discussed above. In addition, the LMS 64 responsible for the upstream LDP Session to node A forwards a label mapping message to it with the label previously reserved by the label manager 62.

Note that for source-routed SLSPs it may not be necessary for the transit node B to obtain next hop information from IP Routing 70. This is, however, a preferred feature which enables the transit node to confirm through its internal routing tables that the next hop provided in the source route list is accurate (e.g., by checking whether the next hop is listed under the requested IP destination address or prefix). If the explicitly routed next hop cannot be confirmed, then an error can be declared.

8.3 *Procedures on Egress Node*

On the egress node C, a label request message is received on the upstream signalling link with node B and forwarded by the L2 processor 50 to the responsible LMS 64. The LMS 64 requests next hop information from MPLS Routing 70 which, in turn, requests next hop information from IP Routing 68. In this case, however, one of the following circumstances arises: (1) the next hop router ID returned by IP Routing 68 is the current node; or (2) the next hop is found, but no LDP Session exists to the next hop (i.e., the edge of the MPLS domain is reached). In either of these cases, MPLS Routing 70 informs the LMS 64 that the SLSP for FEC Z must egress at this node, whereby the LMS 64 sends a label mapping message to the upstream node B as previously described but does not (and cannot) progress the label request message for FEC Z forward. In this case, MPLS Routing 70 adds an entry 126 in the RWT 120, as previously discussed, but also adds the newly created RWT entry 126 to the LET 130.

In this case, the LMS 64 instructs the call processor 72 to establish an SI configured for IP forwarding. This SI has an ATM endpoint (i.e., VPI/VCI) equal to the VPI/VCI used as the MPLS label between nodes B and C for the SLSP.

9. Switching/Routing Activity

Having described the set up of an SLSP for FEC Z, the manner in which IP packets associated with FEC Z are processed is now briefly described. At the ingress node A the IP packets arrive at port 14A1 in the form of plural ATM cells which the IP forwarder 22A reassembles into constituent IP packets. Once the destination IP address of the received packet is known, the IP forwarder 22A examines its forwarding table 30 for the "closest" entry. This will be the entry for FEC Z that was downloaded by IP Routing 68 in connection with the establishment of the SLSP for FEC Z. Thus, the forwarding table 30 provides the egress interface index 76, comprising the identity or address of the egress line card 12B, egress port 14B1 and egress SI number. The egress interface index is attached to the packet. The ingress IP forwarder 22A also selects a TI 18 to transport the packet over the switching fabric 20 to the egress IP forwarder 22B, based in part on the COS field encapsulated in the packet. The packet is then re-segmented for transport across the switching fabric 20 on the selected TI 18 and received by the

egress IP forwarder 22B. The egress IP forwarder 22B, in turn, extracts the egress SI and COS information attached to the packet and modifies it to match the encapsulation indicated by the egress interface index (i.e., egress SI). This includes attaching the VPI/VCI label to the packet. The packet is subsequently segmented into constituent ATM cells and transmitted out of the egress port 14B1 with the VPI/VCI values indicated by the egress SI.

On the transit node B, the ATM cells corresponding to the IP packets are received by an ingress port. The CAM 15 returns an ATM egress connection index, whereby the cells are processed as ATM cells. The ingress line card 12A also attaches internal addressing information retrieved from the CAM 15A to each cell, thereby enabling the cells to be routed to the egress line card which replaces the VPI/VCI value of the cells. The egress line card then transmits the cell using the new VPI/VCI value. Note that in this case the IP forwarding modules 22 were not involved in the switching activity and there was no need to re-assemble and re-segment the IP packet, or perform the IP routing lookup.

On the egress node C, the ATM cells corresponding to the IP packets are received by an ingress port and processed in accordance with the SI configured for the VPI/VCI carried by the cells. This SI is configured such that the cells are sent to the IP forwarding module 22A for re-assembly into the higher-layer IP packets and thereafter processed as regular IP packets.

10. Network Topology Changes

10.1 *New LDP Session*

When a new LDP Session is established on node 10, the LMS 64 signals MPLS Routing 70 about this event and informs it about the interface index for the new LDP Session. The signal arises whether the node is the initiator of the new LDP Session or the respondent. Referring additionally to the flow chart of Fig. 11, at step 190 MPLS Routing searches for the peer router ID IP address in the LSLT 100. If an LSLT entry 194 for this router is found, then at step 192 MPLS Routing 70 examines the

corresponding LDP Session list 106 to ensure that no entries 108 exist for an LDP Session having the same interface index as the new LDP session. If no such entry is found, a new entry 108 is created at step 195. If such an entry is found, an error is returned. If no LSLT entry 104 is found which matches the peer router ID for the newly configured LDP Session, then at step 194 MPLS Routing creates and inserts a new LSLT entry 104, following which the LDP session list entry 106 is created at step 195.

At step 196, MPLS Routing 70 traverses the LET 130. For each RWT entry 126 belonging to the LET, the corresponding FEC is determined from hash table 122, and at step 200 the next hop router ID for that FEC is requested from IP Routing 68. At step 201 the next hop router ID is compared against the peer router ID of the newly configured LDP Session. If no match is found, control returns to step 198, and if a match is found, control passes to step 202. At step 202, MPLS Routing 70 instructs the LMS 64 to send a label request message to the newly reachable peer router for the identified FEC.

10.2 *Signaling Link Failure*

When an LDP Session fails on a node it stops forwarding all SLSPs using the associated VPI/VCI range (stored in the label manager 62) and removes the cross-connects from the node. The node also sends a label withdraw message to the upstream peer for each SLSP associated with the failed LDP Session. For instance, if the MPLS link 84BC (Fig. 7) fails, node B sends a label withdraw regarding FEC Z to the ingress node A. When the label withdraw message is received at the ingress node A, it stops using the path (IP hop-by-hop forwarding is used instead) and immediately re-initiates the steps described previously to re-establish a path for FEC Z. If this does not succeed, then the SLSP for FEC Z is placed on the RTL 116 following which the retry procedures as previously described are effected.

Furthermore, when an LDP Session becomes inoperative in the ingress node A for whatever reason, the LMS 64 informs MPLS Routing 70. As part of this call, the LMS 64 provides MPLS Routing 70 with the peer router ID IP address. MPLS Routing 70 then searches for the peer IP address in the router ID field 104a of the LSLT

100. If there is no entry for the peer IP address, an error is returned. If there is an entry 104a for the peer IP address, the corresponding session list 106 is searched for the failed LDP Session. If there is a matching LDP Session entry 108, it is removed from the session list 106.

The *fit_list_entry pointer 108b of the removed session list entry 106 points to the list of all FIT entries 112 representing all ingress SLSPs using the failed LDP Session. For each of these entries, MPLS Routing 70 immediately tries to re-establish the ingress SLSP as described above to see if there is an alternate LDP Session that may be used to set up the ingress SLSP. If the retry is unsuccessful, the ingress SLSP goes on the RTL 116 and the retry procedures outline above are followed.

10.3 *IP Routing Changes*

Over the course of time, IP Routing 68 may discover a new next hop for FEC Z. For example, in the reference network IP Routing on node B may discover that the next hop for FEC Z should be node D (not shown). Upon such a discovery, IP Routing 68 on node B informs MPLS Routing 70 of the new next hop router ID for FEC Z. MPLS Routing 70 uses the following process to re-route the SLSP for FEC Z: First, it searches for a RWT entry 122 matching the IP prefix address, e.g., FEC Z, which has changed in the IP Routing table 75. In the event no entry 122 is found MPLS Routing returns otherwise it continues and next searches for an LSLT entry 104 that matches the router ID of the new next hop D. If there is an LSLT entry 104 and hence LDP Session to the new router D, MPLS Routing requests the LMS 64 to progress each transit SLSP in the RWT list 124 pointed to by the matching RWT entry 122 using the LDP Session to router D. Thus transit SLSPs are re-routed to the new next hop router D. However, if there is no LSLT entry 102 for the router ID of the new next hop and hence no LDP Session therefor, then MPLS Routing 70 places each transit SLSP in the RWT list 124 corresponding to the old-hop router on the LET 130 and informs the LMS 64 that it should consider such SLSPs as egress SLPs. The LMS 64, in turn, instructs the call processor 72 to set up egress SIs for the egress SLSPs.

MPLS Routing 70 also searches for a FIT entry 112 which matches the affected FEC. If there is a FIT entry 112 that matches the FEC and the ingress_setup flag of the fec-status field 112f is non zero (i.e., the path is set up), MPLS Routing 70 requests that the LMS 64 close the ingress SLSP by sending a label release message to the downstream routers. MPLS Routing 70 then searches for an LSLT entry 104a that matches the router ID IP address for the new next hop. If there is such an LSLT entry, then an LDP Session is selected from the corresponding LDP session list 106, and the procedures for establishing an ingress SLSP are followed as described above.

10.4 *Physical Link Failure*

When a physical link between two nodes fail, then signaling links 82 and 84 (see Fig. 7) for both MPLS signaling and IP routing fail. In the present embodiment, IP Routing 68 realizes that the link is down and updates its routing table 75 before the LMS 64 realizes that any LDP Sessions thereover are down. This is accomplished by suitably setting "time out" periods for LDP Sessions and signaling sessions in IP Routing such that interface failures are reflected much quicker into IP Routing 68 than MPLS Routing 70. Accordingly, IP Routing 68 informs MPLS Routing 70 about a new next hop router ID for affected SLSPs and as previously described MPLS Routing 70 will reroute these SLSP paths from the current node, using the new next hop router identified by IP Routing 68. This is more efficient than tearing down the affected SLSPs back to the ingress node and resignaling them as would have occurred if MPLS Routing 70 realizes the signaling link is down.

The foregoing embodiment has been described with a certain degree of particularity for the purposes of description. Those skilled in the art will understand that numerous variations and modifications may be made to the embodiments disclosed herein without departing from the spirit and scope of the invention.

4. Brief Description of Drawings

Fig. 1 is a system block diagram of a network node which processes ATM cells and IP packets.

Fig. 2 is process flow diagram showing how IP packets are processed in the node of Fig. 1.

Fig. 3 is a diagram of a forwarding table employed by IP forwarders associated with input/output controllers of the node of Figure 1.

Fig. 4 is a diagram of a data structure representing a "service interface" associated with nodes such as shown in Fig. 1.

Fig. 5 is an architectural block diagram of hardware processors and software processes associated with a control card on the node of Fig. 1.

Fig. 6 is a master IP routing table associated with an IP network.

Fig. 7 is a diagram of a reference network illustrating an MPLS domain within an IP network.

Fig. 8 is a schematic diagram of a database employed by the node of Fig. 1 to manage signalled label switched paths (SLSPs).

Figs. 8A and 8B show certain fields of the database of Fig. 8 in greater detail.

Figs. 9 and 10 are logic flow charts showing the steps executed by the node of Fig. 1 in establishing an SLSP.

Fig. 11 is a logic flow chart showing the steps executed by the node in the event a new SLSP signalling link is established.

Fig. 1

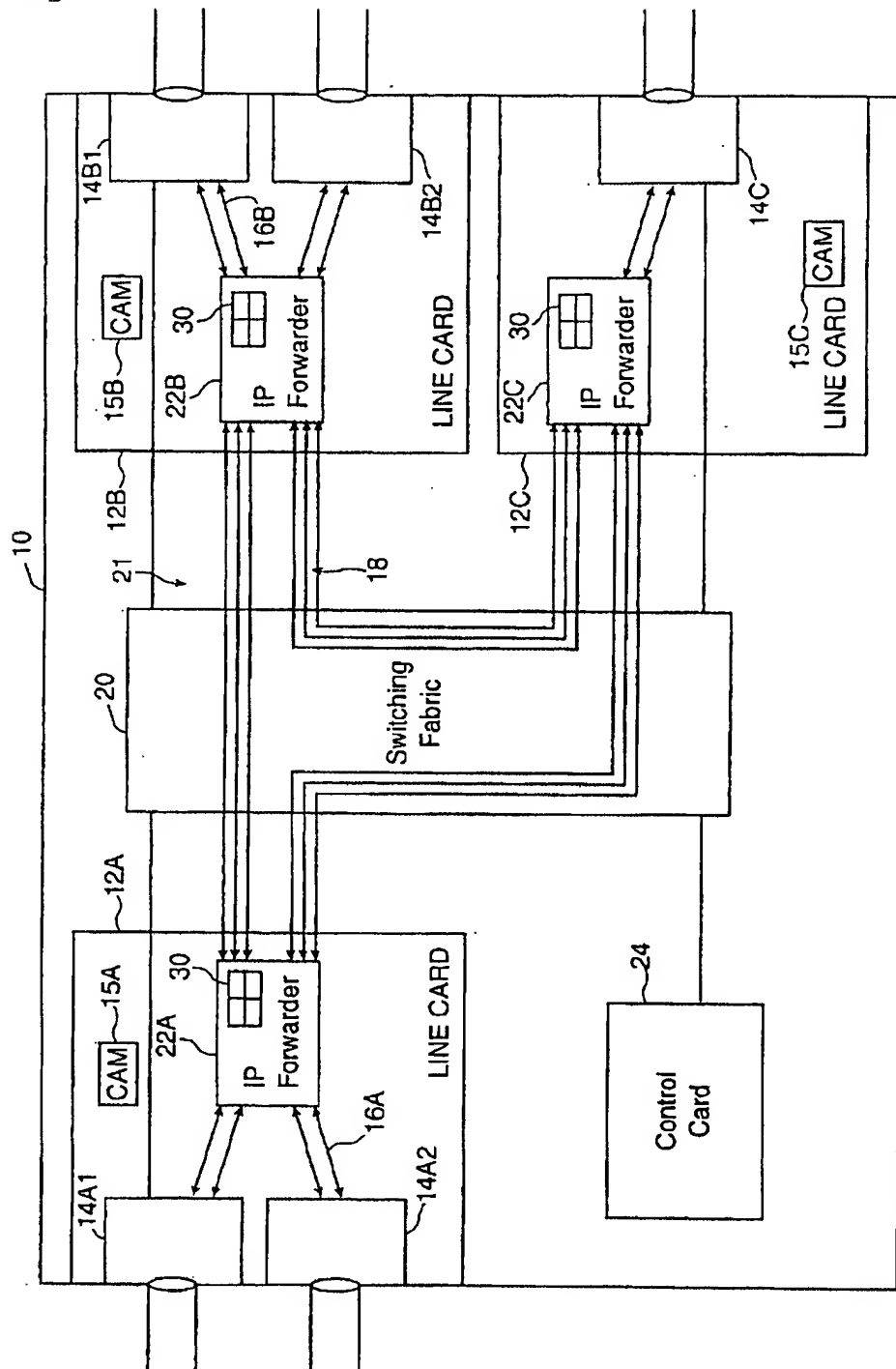


Figure 1

Fig. 2

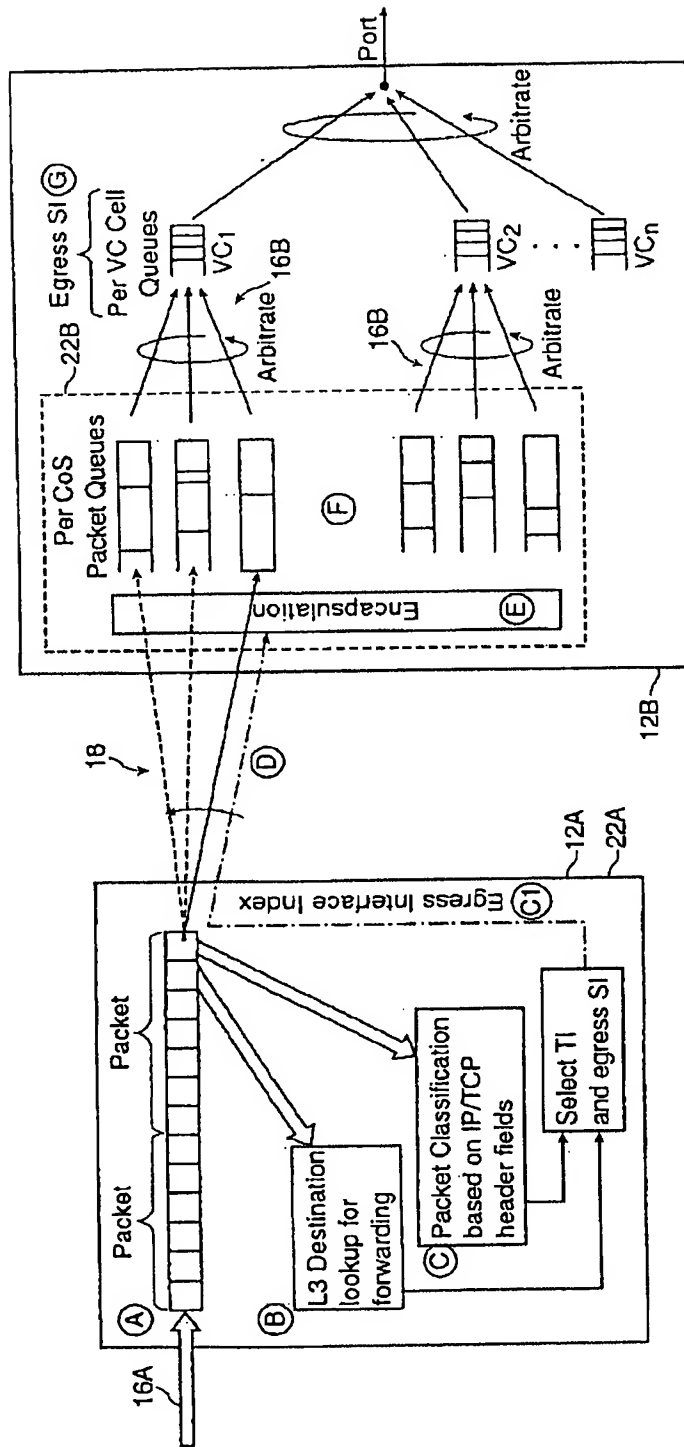


Figure 2

Fig. 3

IP Address	Egress Interface Index
FEC Z → 1.2.3.4	76

30

Figure 3

Fig. 4

Fig. 4

Parameter	Description	Default	NMTI	SNMP	NCI	CLI
ID number	A identification number assigned for this SI; unique within a subslot. This value is internally assigned and cannot be changed. This is 5 digit number field.	None	R	-	-	-
Endpoint	The ATM endpoint (shelf-slot-subslot-port; VPI/VCI) used by the SI.	None	R/W	-	-	-
Name	Name of the SI. This is 16 character text string field	Empty	R/W	-	-	-
Application	Application(s) provided by this SI. This is a boolean vector (i.e. bit map) indicating whether each of forwarding, routing and LDP is enabled.	Forward	R/W	-	-	-
Address type	Type of the IP address field. Valid type supported are unnumbered and IPv4.	Un-numbered	R/W	-	-	-
IP address	The IP address of the service interface. Represented to the user in standard "dotted decimal" format. "Illegal" IP addresses (e.g. 0.0.0.0, 255.255.255.255) are blocked.	Un-assigned	R/W	-	-	-
IP address prefix length	Number of bits in the IP address which constitute the (sub)network ID. A number in the range of 0..32.	None	R/W	-	-	-
Neighbour address type	Type of the neighbour IP address field. Valid type supported are unnumbered and IPv4.	Un-numbered	R/W	-	-	-
Neighbour IP address	The IP address used at the termination of the SI at the neighbouring router. Represented to the user in standard "dotted decimal" format. "Illegal" IP addresses (e.g. 0.0.0.0, 255.255.255.255) are blocked.	Un-assigned	R/W	-	-	-
Encapsulation	Encapsulation used on the SI. (RFC1483 LLC/SNAP routed IP, RFC1483 NULL)	RFC1483 NULL	R/W	-	-	-
MTU	Maximum Transmission Unit.	2016 octets	R	-	-	-
Ingress traffic contracts	An ingress traffic contract structure consists of an action (disable, tag, discard), a committed information rate (in b/s) and a burst size (in bytes). Eight ingress traffic contract structures are contained in each SI; each applies to a CoS.	disable CIR 0 BS 0	R/W	-	-	-
Status	Status of the service interface. (Up, Down).	Down	R	-	-	-

Service Interface Parameters

Fig. 5

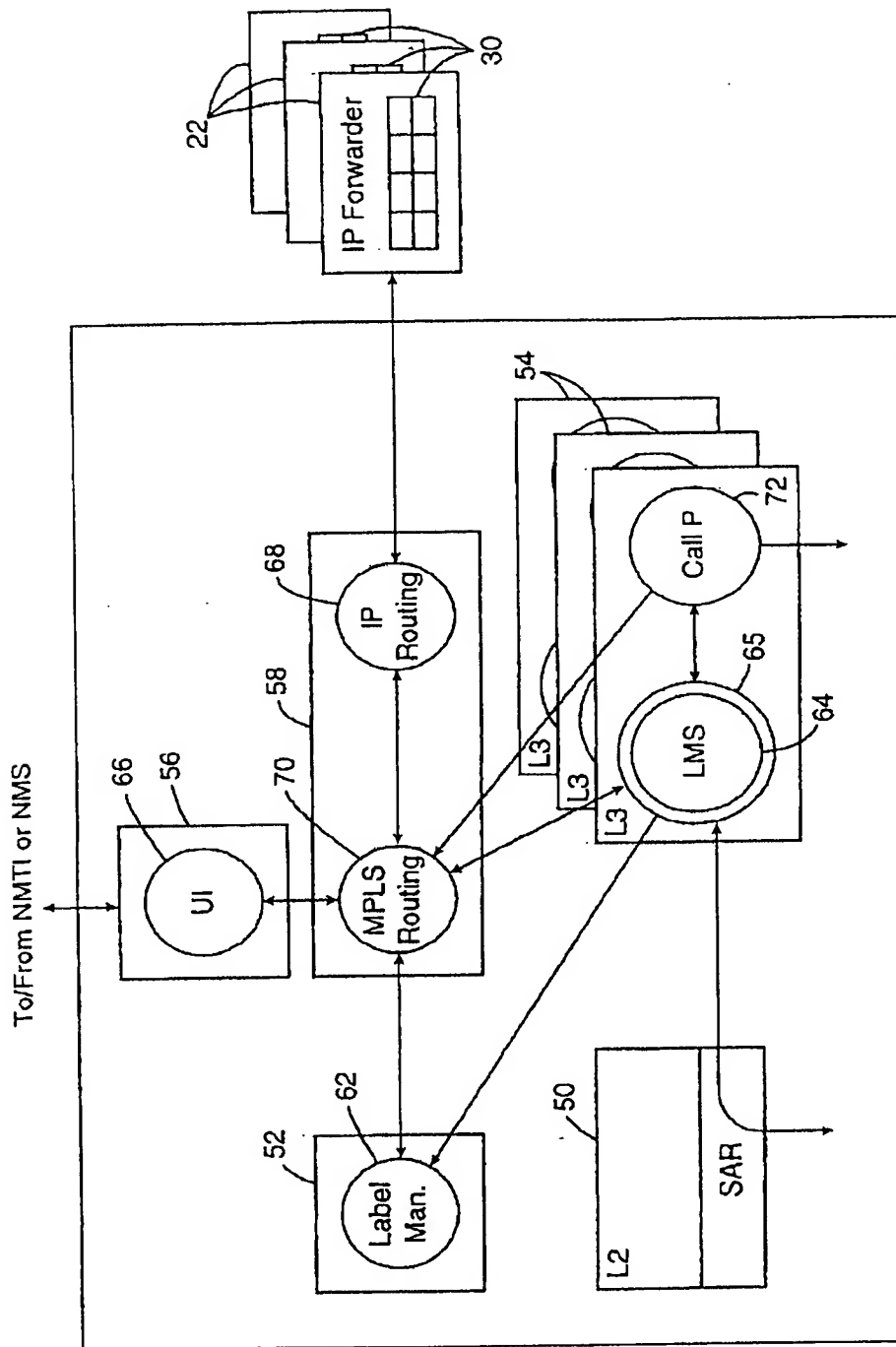


Figure 5

Fig. 8A

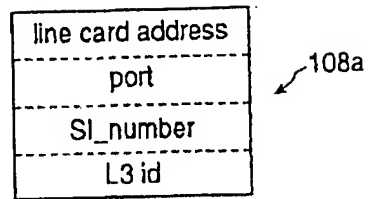


Figure 8A

Fig. 8B

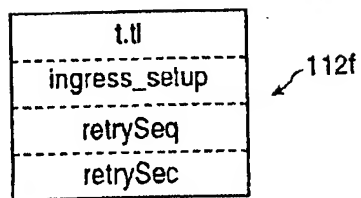


Figure 8B

Fig. 9

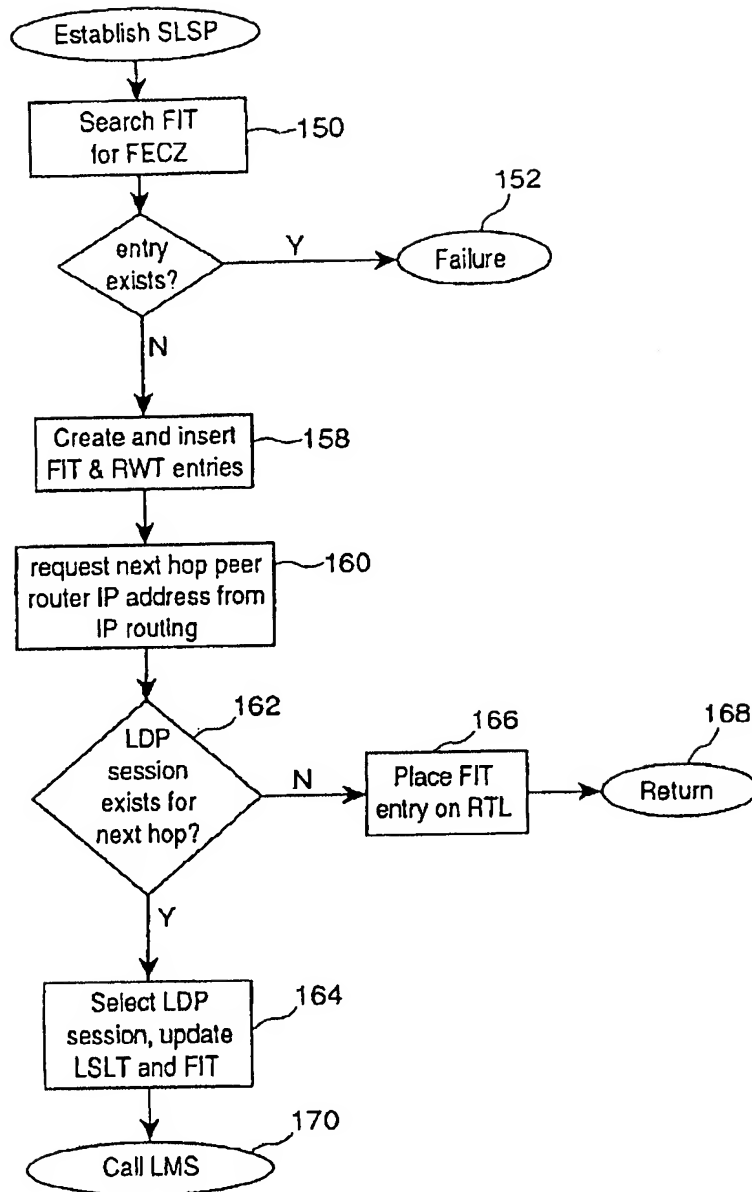


Figure 9

Fig. 10

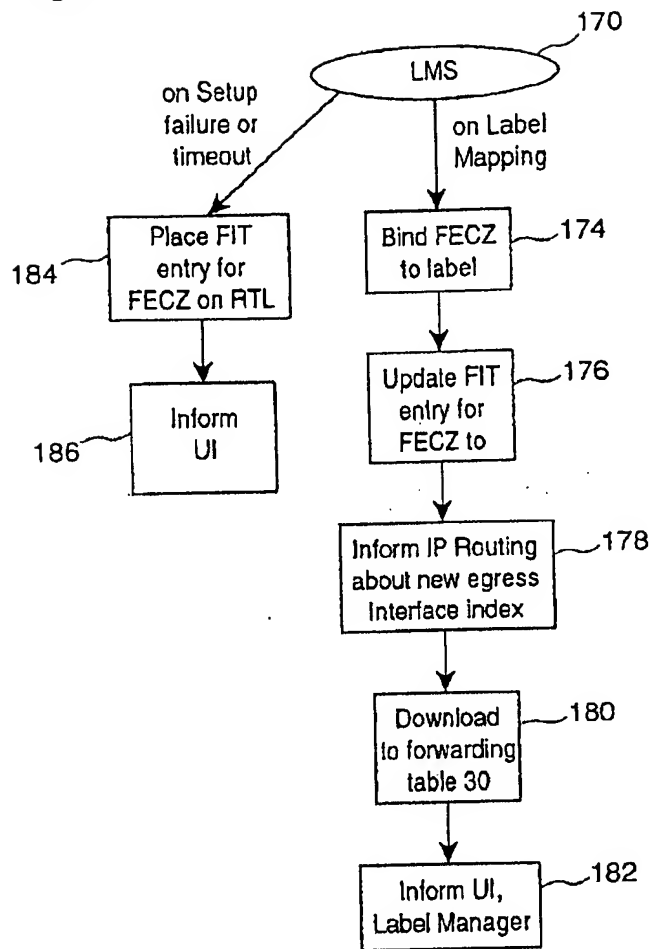


Figure 10

Fig. 11

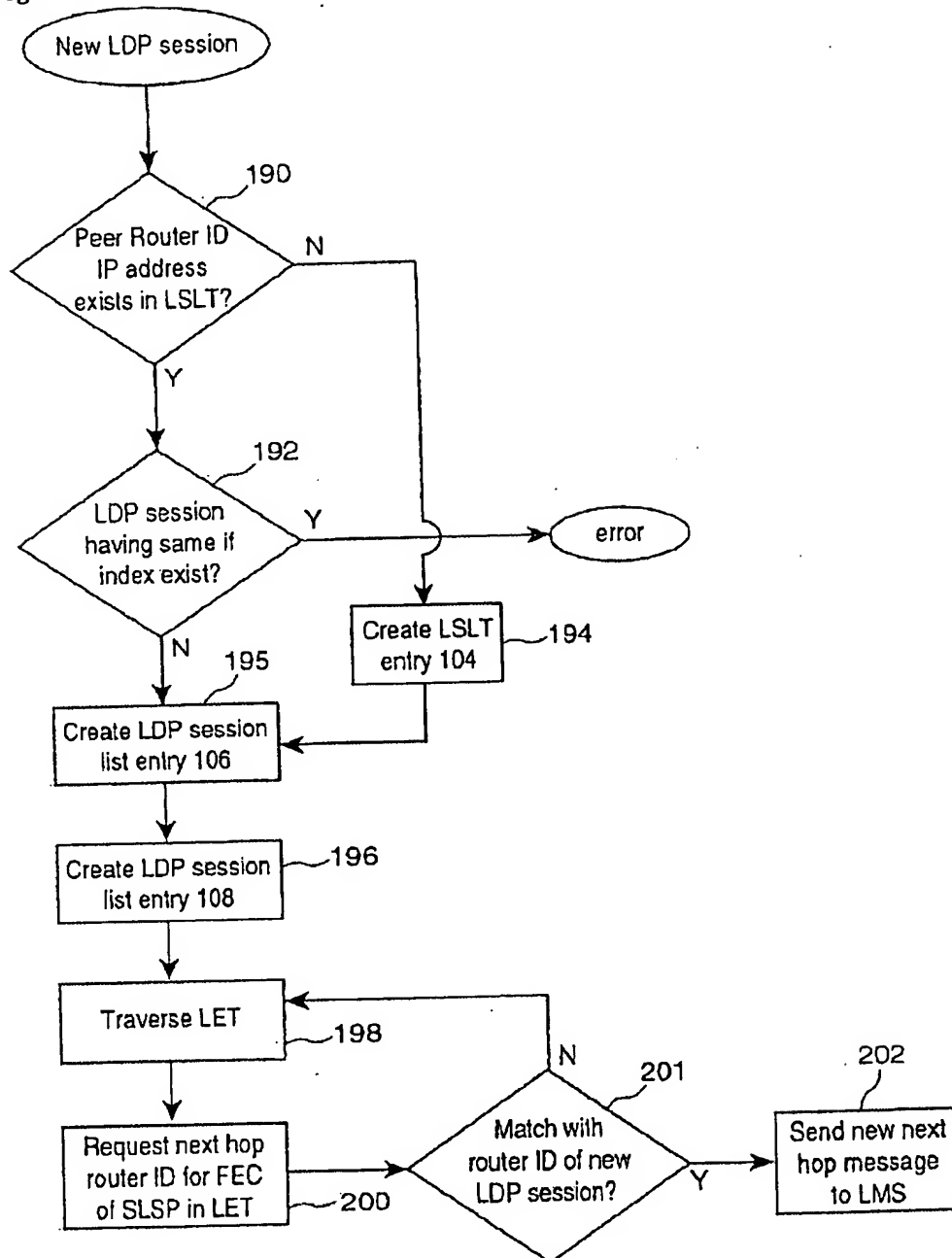


Figure 11

1. Abstract

A method of timing an attempt to establish a connection path between a first and second node in a communications network is provided. The method initiates the attempt to establish a connection path after a period of time has elapsed wherein the period of time is greater than another period of time which had previously elapsed between two previous attempts, if any, to establish the connection.

2. Representative Drawing

Fig. 1